DOCUMENT RESUME

ED 293 877 TM 011 467

AUTHOR Arter, Judith A.; Salmon, Jennifer R.

TITLE Assessing Higher Order Thinking Skills. A Consumer's Guide.

INSTITUTION Northwest Regional Educational Lab., Portland, Oreg.

SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Apr 87
CONTRACT 400-86-0006
NOTE 78p.
PUB TYPE Guides - Non-Classroom Use (055) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Cognitive Tests; *Critical Thinking; Decision Making Skills; Educational Assessment; Elementary Secondary Education; Problem Solving; Test Reliability; Test Reviews; *Test Selection; Test Use; Test Validity
IDENTIFIERS *Higher Order Skills; *Thinking Skills

ABSTRACT

Information necessary for informed use by practitioners of higher-order thinking skills (HOTS) tests is presented. Assessment instruments that (1) have as a basic assumption that thinking skills can be taught; (2) provide information that can be useful for instructional planning related to complex thought processes; and (3) have content that is taught, provide information that can be useful for instructional planning related to complex thought processes, and have content that is related to skills needed to function with reason in the real world are discussed. Content, format, and purpose must be considered in test selection. Although most current tests are designed for Grades 4 and above, some are designed for grades as low as pre-Kindergarten. Most tests cover a broad grade span. Cognitive skills included in HOTS are discussed, assessment issues are reviewed, the current state-of-the-art in assessment is outlined, and steps in selecting a HOTS test are described. Appendix A contains individual reviews of 10 HOTS tests and brief descriptions of other critical thinking, problem solving, or decision making tests. Two developmental tests are reviewed; eight others are briefly described, as are specific tests of creativity, achievement, and ability. Local, state, and federal programs with testing resources are described in other appendices. Appendix D contains a checklist for selecting HOTS tests. Tests are further indexed by grade level and title. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

+ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jerry D. Kirkpatrick

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Assessing Higher Order Thinking Skills

A Consumer's Guide

By

Judith A. Arter

and

Jennifer R. Salmon

April 1987

Northwest Regional Educational Laboratory
Evaluation and Assessment
101 SW Main Street, Suite 500
Portland, Oregon 97204
(503) 275-9500

April 1987

This guide is published by the Test Center of the Northwest Regional Educational Laboratory, a private nonprofit corporation. The work contained herein has been developed under Contract 400-86-0006 for the Office of Educational Research and Improvement (OERI), U.S. Education Department. The opinions expressed in this guide do not necessarily reflect the views of OERI and no official endorsement by the office should be inferred.

TABLE OF CONTENTS

1. Purpose For The Consumer's Guide 1
2. The Importance of Looking at Higher Order Thinking Skills 1
3. Definitions 1
4. Assessment Issues 4
5. State-of-the-Art 7
6. Futures 9
7. How To Select a Higher Order Thinking Skills Test 9
8. References 11

Appendix A: Reviews of Higher Order Thinking Skills Tests 15

- Critical Thinking, Problem Solving or Decision Making Tests 17
- Other Critical Thinking, Problem Solving or Decision Making Tests 39
- Developmental Tests 41
- Other Developmental Tests 45
- Creativity Tests 47
- Achievement Tests 49
- Ability Tests 51
- Observation Scales 51

Appendix B: Summary Table of Higher Order Thining Skills Test 53

Appendix C: Testing Resources 65

- Local, State and Federal Department of Education Programs 67
- Colleges, Universities, and Professional Organizations 71
- Bibliographies of Higher Order Thinking Skills Assessment Tools 71
- Books and Articles 73
- Newsletters 73
- Curriculum Reviews 75

Appendix D: Checklist for Selecting a Higher Order Thinking Skills Test 77

- Checklist for Selecting a Higher Order Thinking Skills Test 79

Title Index 83

Index of Tests by Grade Level 85

1. Purpose For The Consumer's Guide

There is currently a great deal of interest in placing more emphasis on higher order thinking skills in the schools. Some authors (for example, Kearney et al, 1985, p. 49) claim that interest is greater now than at any time in the past. Schools are interested in finding out how well their students think and in improving that thinking ability. This Guide is intended to assist this process by providing an overview of the current state-of-the-art in assessing higher order thinking skills (HOTS).

This Guide is intended for use by practitioners. It is intended to provide the information necessary for users to become more informed and thoughtful consumers of HOTS tests. Included in the Guide are a brief discussion of the issues in assessing HOTS, reviews of over 40 tests and other assessment devices, guidance on how to select a test of higher order thinking, and a listing of other resources for those interested in pursuing the topic further.

2. The Importance of Looking at Higher Order Thinking Skills

The following reasons for increased emphasis on assessing and teaching higher order thinking skills have been suggested by recent authors:

1. There is evidence that good thinking is not widespread (Norris, 1985; Walsh, 1985; Kearney et al, 1985). One source of evidence for this is decreasing test scores in the upper quartile of students (Reidman, 1985; Sternberg, 1984b).
2. We need the ability to judge, analyze and think critically in order to function in a technological society and in a democracy (Reidman, 1985; Kneedler, 1985; Kearney et al, 1985).
3. Such skills can be taught and all students can improve in their ability to think (Sternberg, 1984b, 1985, 1986; Costa, 1983; Lipman, 1985; Baron and Kallick, 1985; Kneedler, 1985).
4. Assessing HOTS can provide impetus for driving the curriculum (Reidman, 1985; Kearney et al, 1985).

3. Definitions

Players

Assessment of intelligent behavior is the concern of those interested in intelligence testing, Guilford's Structure of the Intellect, developmental models (e.g., Piaget), critical thinking, creativity, problem solving, achievement testing, and curriculum development. Given all these players, when looking for tests of higher order thinking skills, the first question is what will be included? Sternberg (1986) outlines three major theoretical approaches to intelligence. Psychometric researchers are interested in the structure of the mental abilities that constitute intelligence. Guilford, for example, attempted to delineate all possible thinking skills (called mental abilities). A Piagetian approach seeks to understand the stages in the development of intelligence. This perspective examines how intelligence develops rather than looking at its structure. Finally, cognitive researchers are interested in the processes of intelligence. They seek to understand the ways that people mentally represent and process information in order to respond to various tasks. People examining higher order processing such as reasoning and problem solving fall into this category; but so do those examining any task requiring thinking, however abstract (for example, how we process analogies).

We believe that when educators discuss higher order thinking skills they mean something different from all possible thinking skills as outlined in the above approaches. They are specifically focusing on complex thought processes required to solve problems and make decisions in everyday life, and those

that have a direct relevance to instruction. Therefore, we constrain our reporting of assessment instruments to those which:

1. have as a basic assumption that thinking skills can be taught;
2. provide information that can be useful for instructional planning related to complex thought processes; and
3. have content that is related to skills needed to function with reason in the real world.

Because of this practical approach, the assessment devices related to critical thinking and problem solving appear to be most relevant. Therefore, this type of instrument will be emphasized most in this review in terms of definitions, assessment issues, and state-of-the-art. This type of instrument will also be emphasized in the long reviews in Appendix A. However, shorter reviews of other types of instruments will be included because of their relevance and potential usefulness for instruction and evaluating programs.

Intelligence (Ability) Tests. Sternberg (1984a) points out that intelligence tests do measure some components of intelligent behavior, not so much because the items represent tasks needed in everyday life but because the metacognitive and performance skills required to think through a problem in everyday life are those brought to bear to answer the questions on the test. DeBono (1977) agrees--"IQ tests manifestly require the exercise of thinking. But IQ tests are not a test of thinking" (p. 225). Intelligence tests might, therefore, be used to measure the outcomes of a curriculum in HOTS, but most are not useful for the purposes we have outlined above--they report a single score which is not relevant to the skills we are considering, they imply that intelligence is fixed and innate, and they often have very abstract item types.

There are some intelligence or ability tests which might provide information useful for instruction. For example, those tests based on Guilford's Structure of the Intellect may be useful because they outline the basic processes on which the more complex processes, such as critical thinking, rely (Presseisen, 1985). We have included brief descriptions of a few of these tests.

Creativity. Creativity is included in many concepts of HOTS. This field is, however, very large. It is outside the scope of this report to include all the measures and assessment issues involved with assessing creativity. Although not intended to be comprehensive, some of these instruments are included in this review.

Developmental Approaches. There are instructional materials and tests based on developmental theories (e.g., Piaget). These claim to assist in furthering development toward formal reasoning. Therefore, some of these instruments are included.

Achievement Tests. Achievement tests have always included items going beyond recall. Examples are math problem solving, making inferences from graphs and charts, and making inferences from reading passages. Recently many publishers either provide separate subtests to measure these areas (e.g., CIRCUS Think It Through) or rescore the existing items to provide a HOTS score (e.g., Metropolitan Achievement Test, 6th edition). Therefore, we have included short reviews of current major achievement test series which provide information on HOTS.

The reviews of assessment devices provided in Appendix A are organized by the categories outlined above--problem solving/critical thinking, developmental, creativity, achievement and ability tests.

What Constitutes Higher Order Thinking Skills?

There have been differences between the skills included in the concept of HOTS depending on the perspective of the author. Sternberg (1985) and Quellmalz (1985) outline three previously independent approaches to the topic. The philosophers concentrated on the assessment of authenticity, accuracy and worth of knowledge claims and arguments (Beyer, 1985). This was generally called critical thinking and included such things as formal logic, judging the credibility of a source of information, and

discovering flaws in arguments (Quellmalz, 1985; Presseisen, 1986). Psychologists identified reasoning skills and their underlying cognitive processes. Finally, educators looked at classes of tasks, leaning heavily on the upper categories of Bloom's taxonomy--analysis, synthesis, comparison, inference and evaluation.

Recently, several authors have tried to consolidate the various conceptions of HOTS to provide an overall picture of the skills involved. This has been part of a general movement of both philosophers and psychologists to join forces (Presseisen, 1986). Ennis (1987) and Gubbins (as reported in Sternberg, 1985) provide two good summaries of these skills.

Gubbins' matrix of thinking skills is presented in Figure 1. In a sense, Figure 1 is a definition of what we call HOTS because it lists the skills considered by a consensus of authors to be components of that concept. We will use the term "higher order thinking skills" to refer to this entire constellation of skills to avoid the impression that we are dealing with any single theoretical approach.

I. Problem Solving
- A. Identifying general problem
- B. Clarifying problem
- C. Formulating hypothesis
- D. Formulating appropriate questions
- E. Generating related ideas
- F. Formulating alternative solutions
- G. Choosing best solution
- H. Applying the solution
- I. Monitoring acceptance of the solution
- J. Drawing conclusions

II. Decision Making
- A. Stating desired goal/condition
- B. Stating obstacles to goal/condition
- C. Identifying alternatives
- D. Examining alternatives
- E. Ranking alternatives
- F. Choosing best alternative
- G. Evaluating actions

III. Inferences
- A. Inductive thinking skills
  1. Determining cause and effect
  2. Analyzing open-ended problems
  3. Reasoning by analogy
  4. Making inferences
  5. Determining relevant information
  6. Recognizing relationships
  7. Solving insight problems
- B. Deductive thinking skills
  1. Using logic
  2. Spotting contradictory statements
  3. Analyzing syllogisms
  4. Solving spatial problems

IV. Divergent Thinking Skills
- A. Listing attributes of objects/situation
- B. Generating multiple ideas (fluency)
- C. Generating different ideas (flexibility)
- D. Generating unique ideas (originality)
- E. Generating detailed ideas (elaboration)
- F. Synthesizing information

V. Evaluative Thinking Skills
- A. Distinguishing between facts and opinions
- B. Judging credibility of a source
- C. Observing and judging observation reports
- D. Identifying central issues and problems
- E. Recognizing underlying assumptions
- F. Detecting bias, stereotypes, cliches
- G. Recognizing loaded language
- H. Evaluating hypotheses
- I. Classifying data
- J. Predicting consequences
- K. Demonstrating sequential synthesis of information
- L. Planning alternative strategies
- M. Recognizing inconsistencies in information
- N. Identifying stated and unstated reasons
- O. Comparing similarities and differences
- P. Evaluating arguments

V. Philosophy and Reasoning
- A. Using dialogical/dialectical approaches

Figure 1
Gubbins' Matrix of Thinking Skills

Note: This matrix is based on a compilation and distillation of ideas from Bloom, Bransford, Bruner, Carpenter, Dewey, Ennis, Feuerstein, Jones, Kurfman and Solomon, Lipman, Orlandi, Parpes, Paul, Perkins, Renzulli, Sternberg, Suchman, Taba, Torrance, Upton, the Ross Test, the Whimbey Analytical Skills Test, The Cornell Critical Thinking Test, the Cognitive Abilities Test, the Watson-Glaser Critical Thinking Appraisal, the New Jersey Test of Reasoning Skills and the SEA Test.

Definitions of the subcomponents in Figure 1 are likewise defined by the skills they subsume. Problem Solving refers to thinking processes used to resolve a known or defined difficulty. Decision Making refers to using basic thinking processes to choose a best response among several options; that is, deciding the best way to go about doing something. Inferences and Evaluative Thinking Skills refers to that cluster of skills traditionally considered to be Critical Thinking (Beyer, 1985)--evaluating the reasonableness of arguments intended to persuade one to have a certain opinion. Divergent Thinking Skills refers to one component of creativity. Creativity is using basic thinking processes to develop or invent novel, aesthetic, constructive ideas or products (Presseisen, 1985). Divergent thinking is the ability to produce a lot of unusual ideas.

Metacognition. Metacognition is that set of executive processes which decides which strategy to use to solve a problem and monitors how the strategy is working to solve the problem (Sternberg, 1986, p. 17). Some of these skills are included in Gubbins' Matrix. For example, "monitoring acceptance of the solution" and "deciding on the nature of the problem." Others may be implied but are not directly stated, for example "being sensitive to external feedback" and "allocating resources for problem solution." There are no assessment instruments which purposefully intend to assess metacomponents. However, many of these skills are implied or included on instruments assessing other HOTS.

Affect. So far we have discussed only the cognitive skills included in HOTS. In order to be a good reasoner in the real world, not only do we need cognitive skills, but we also need the disposition or motivation to reason (Ennis, 1987; Sternberg, 1986; Passow, 1981, Norris, 1985). This is similar to the issue of children who can read but don't, because they are not motivated to do so. Ennis (1987) sees dispositions as including a person's willingness to be open-minded, well-informed, change positions when the evidence warrants, stick to the point, and be sensitive to the feelings, level of knowledge and degree of sophistication of others. Similarly, Paul (1986) considers "strong sense critical thinking" to include the willingness not only to reason but also to examine one's whole frame of reference and belief system. These affective dispositions are not specifically included in the taxonomy we use to classify instruments. We have found a few instruments that look at dispositions.

4. Assessment Issues

Structured Format

Most of the instruments reviewed had structured formats--multiple choice, matching, etc. Such questions require that only one answer be correct. A problem arises in that most definitions of HOTS include the ability to think through real-world problems which typically lack a clear formulation, a procedure that guarantees a correct solution, and criteria for evaluating solutions (Fredericksen, 1984; Paul, 1986). There are often multiple correct solutions to this type of "fuzzy" problem because more than one answer could have a defensible rationale for choice. Therefore, the situations that we are most interested in assessing are those that are most difficult to put in structured format because of the requirement to have one right answer.

For example, the following test question from the **Cornell Critical Thinking Test, Form X** (1985) could have a different, but equally correct answer depending on one's level of knowledge, sophistication or cultural background.

> The test taker is to imagine he or she is part of a second group of explorers to land in Nicoma. In one part of the test, the exercise is to decide which of two statements is more believable, or if the statements are equally believable.

27. A. The health officer says, "This water is safe to drink."

B. Several others are soldiers. One of them says, "This water supply is not safe."

C. A and B are equally believable.

The keyed answer is A, the health officer's statement is more believable. For children raised in military families or who know that a soldier is trained in outdoor survival, the correct answer is B. For children of cultures who are raised to distrust government officers, the correct answer may be B or perhaps C.

Process Versus Single Solution. Some people see structured-format tests as stressing getting the right answer rather than stressing either the process by which the answer is obtained or the person's ability to defend their answer (Norris, 1985; Costa, 1983). They argue that the ability to come up with a position and defend it is a truer measure of HOTS specifically because a person's philosophical orientation and culture could lead to alternative "correct" positions (McPeck, 1981). Thus, good HOTS test situations would preclude having only one right answer.

Novelty. HOTS are assessed only when a situation is novel. Otherwise scores are contaminated by level of knowledge of the examinee (Reidman, 1985; Ennis, 1987; Costa, 1983). A problem requiring HOTS for one person may not for another. For example, several items on the **Understanding In Science Test** requires students to predict the course a ball will take after bouncing off a wall based on the angle it hits the wall. For young students this might require HOTS because they must bring past experience and association skills to bear on the solution. For a physics student this may only require recall of information. Thus, items equally measure HOTS only to the extent that they are equally novel to all examinees. Therefore, even formal logic items could be measures of achievement rather than measures of HOTS for those students who have been instructed in formal logic.

Establishing Test Validity. Defenders of structured-format tests claim that a right answer can be used as a proxy for good thinking if it can be demonstrated that good thinking leads to the right answer and faulty thinking leads to wrong answers, and that these patterns hold across groups typically tested in this country. For example, **The Test of Appraising Observations** (Norris and King, 1984) attempted to validate the measure based on whether the responses were based on good or poor thinking.

In general, however, many of the structured-format tests either deal with this issue by resorting to well-structured (and therefore less interesting) item types, or they ignore this issue and put items on their tests which could be influenced by knowledge, sophistication or philosophical orientation of the test taker without proper documentation that this does not occur. In our review we have tried to point out these variations between tests.

Open-Ended Tests

Critics of structured-format tests maintain that the only reasonable way to assess HOTS is to observe or sample actual performance of a task, for example, writing essays to support a point of view. But, there are problems with these approaches also. It is more difficult to score open-ended tests in an objective, uniform manner. For example, the philosophical orientation of the scorer might affect the score given to an examinee if the examinee's position is opposed to that of the scorer.

Creators of open-ended tests must report on the consistency of scores across scorers and how scores are unbiased (Ennis, 1986).

Can HOTS Only Be Measured Within The Context Of Specific Subject Matters?

McPeck (1981) and Reidman (1985) believe that HOTS can only be assessed within a subject matter domain because any meaningful thinking requires information outside the problem situation as posed in the test, and because the types of skills important for one subject domain may not be the same as for another. Presseisen (1985) also comments that "decision making" skills may be more important for social studies and careers, "problem solving" skills may be more important in science and math, "critical thinking" skills may be more important in debate, government and language arts, and "creativity" may be more important in the fine arts.

Other authors disagree (e.g., Glatthorn and Baron, 1985). Many instructional programs emphasize teaching HOTS as a separate subject matter (e.g. Feuerstein and Philosophy for Children). Ennis (1986) and Sternberg (1984b) feel that general principles can be taught as a separate entity, but then students need to be explicitly shown how to apply them to various content areas. Ennis (1987) points out that, obviously, test questions have to be about some topic. But these topics can be drawn from daily life.

Frederickson (1984) implies that when the items on a test are well-structured, problem solving becomes more specific to a particular content area. When one attempts to problem solve in a fuzzy area, the skills brought to bear are more general.

In our reviews we point out assessment devices that are specific to certain content area domains, and those which are intended to be general measures of HOTS in everyday life.

Understanding The Task

The test developer has to make sure that the student understands the task. Unfamiliar vocabulary in instructions or in the problem situation can render an item more a test of vocabulary, reading, listening, or writing than HOTS (Morante and Ulesky, 1984; Ennis, 1987).

Construct Validity

Construct validity means that the test actually measures the underlying concepts that it claims to measure. Sometimes it is difficult to know what evidence would be acceptable evidence that a test measures HOTS. Validity is often shown by correlating scores on the HOTS test with ability or achievement test scores. Should correlation between HOTS scores and ability test scores be high or low? If they are too high then why have a separate HOTS test? They can't be too low either because there must be some connection between the thinking required on the two tests.

Similarly, what should be the correlation between achievement test scores and HOTS test scores? If HOTS test scores are not correlated to achievement then the concept is not useful--one reason we want to improve HOTS scores is to improve achievement. If HOTS test scores are highly related to achievement, then why have a separate HOTS test?

The **Ennis-Weir Essay Test** presents no evidence of validity because the authors claim that no satisfactory criterion has been established. The **Test of Appraising Observations** uses

independent observation of students thinking processes while they are taking the test as a criterion for its validity.

This issue has not been satisfactorily resolved.

Atomistic vs. Holistic Assessment

There is one final criticism of many approaches to assessing HOTS--whether we should try to break down the HOTS concept into subskills and assess each separately or whether the concept of reasoning requires interplay between the components. Moss and Petrosky (1983) and Quellmalz (1985) feel that one cannot define critical thinking as a series of discrete skills or steps because critical thinking skills are interdependent and part of an integrated process. By testing independent skills one loses the whole. The issue here is parallel to that in reading--testing individual skills such as decoding versus testing the ability to read (McPeck, 1981).

This is, of course, also a problem for instruction in HOTS as well as assessment. This issue is still being debated.

5. State-Of-The-Art

Given all the issues and considerations presented in the last section, how do current assessment instruments stack up?

Format

Most current assessment instruments that are readily available to consumers have a structured-format. The Ennis-Weir is an essay test. Several state departments of education and local school districts have also developed essay tests. These are, however, not readily available. Creativity tests are generally open-ended, but there is some concern whether divergent production instruments actually measure creativity (Perkins, 1985).

Grade Levels

Most current tests are designed for grades 4 and above. We found some that were designed for grades as low as preK. Most tests cover a broad grade span which might make them less effective for any single grade.

Content

The assessment devices measure a variety of skills. This reflects the different theoretical approaches taken by the authors as well as differences in definitions within any single tradition.

Item-content varies widely from attempts to present real-world situations to items that are very abstract. The rationale for the former type of item is that they directly measure the skills we want students to have. The rationale for the latter is that they tend to have only one right answer and seem to be related to other items which cluster around a skill domain. Some tests have parts that look like achievement, ability or readiness tests.

Critical thinking, creativity and developmental approaches are pretty well represented. We found fewer tests in problem solving and decision making. Problem solving tests may be more embedded in subject domains.

Most of the assessment instruments reviewed support the idea of general-knowledge HOTS tests. They are not embedded in subject area content.

Most achievement test batteries are overtly including HOTS items on their tests and sometimes report separate subscores for HOTS. These seem to be mainly based on Bloom's taxonomy.

Most tests emphasize testing individual HOTS skills rather than taking a holistic view of the skill domain. Some essay tests are scored both analytically and holistically.

There are very few tests of HOTS dispositions. We found two assessments of creativity disposition.

Validity

Examination of validity is generally pretty weak on many of the tests (Morante and Ulesky, 1984). Two good instruments in this regard are the **Test of Appraising Observations** and the **Cornell Critical Thinking Test**. Many others rely mostly on face/content validity. In addition, if other information is presented it is not explained how the results give evidence of validity.

Criteria for examining the validity of HOTS assessment measures is presented in Appendix D. These criteria relate to the assessment issues presented in Section 4, above.

Reliability

Total score reliability on the tests are generally pretty good--reliabilities are generally above .80. Subcomponent or subtest scores are generally lower. This makes profiling individual students on subcomponent skills problematic.

Usability

Because most instruments have a structured-response format, they are generally easy to use. There are many tests that are professionally packaged, readily available from publishers, relatively inexpensive and machine-scored by the publisher.

There is a certain skimpiness when it comes to assistance with interpreting and using the results. The instruments best in this area are those associated with specific curriculum materials (e.g., the **New Jersey Test of Reasoning Skills** and the **Philosophy for Children Program**). Very few of the tests have norms. Those that do have norms rarely provide norm dates; and the norming sample is often small and restricted.

Summary

Several authors reviewed for this report (including Ennis, 1987) felt that the area of HOTS testing is currently primitive both because of issues having to do with structured format tests and also because of the general lack of formally developed instruments to gather information in other ways (e.g., interviews, essays and observations). It appears that some of the structured format tests are on the right track in terms of focus, relating items to real-life situations and validation. Also more observational and open-ended devices will become available. The tests are not perfect, and there are still some questions as to exactly what some of them measure. However, with care it seems that the assessment devices currently available can provide some useful information about the HOTS abilities of students.

6. Futures

Several authors outlined what the future holds in store for assessing HOTS. Included were:

- There will be more tests developed in the near future (Ennis, 1987).
- Computer simulations will be developed which would more closely approximate certain kinds of performance, for example designing science experiments (Brennan and Stenzel, 1985; Ennis, 1987).
- Tools for assessing dispositions will be produced. For example, the Connecticut Department of Education is working in this area.
- Building alternative responses on structured-format tests by interviewing students will occur more frequently (Norris, 1985).
- There will be more clever ideas on how to make structured-format tests more like open-ended tests; for example, having students choose their answer and then choose the reason for responding like they did (Ennis, 1987).
- There will be more grade-specific tests and more choices in terms of subject matter specificity and particular skill specificity (Ennis, 1987).
- Tests will emphasize real-life situations more heavily.

7. How To Select A HOTS Test

Step 1 -- Decide On Content, Format and Purpose

The first step in choosing an instrument is defining what is locally meant by HOTS. Figure 1 might assist in focusing on the emphasis desired. But also one should consider other approaches such as the developmental and structure of the intellect approaches mentioned in section 3. Based on the information presented in sections 3 through 5 of this report, you also need to determine whether you want an aspect-specific test or a general test of HOTS, a structured-format or open-ended test and whether the test needs to be specific to a particular subject domain.

Caution--Depending on the combination of factors you choose, you might not find exactly what you want. You may also want to base your choice on what is instructionally possible.

Step 2 -- Choose Two or Three Instruments To Review

The instruments in Appendix A are arranged by their major emphasis as outlined in section 3--problem solving/critical thinking, creativity, developmental, achievement and ability tests. In the reviews we have tried to provide information about specific content coverage, type of test and items, reliability, validity and usability so that users can judge which instruments, if any, might satisfy their needs.

Step 3 -- Review The Instruments in Detail

We recommend that users obtain more than one instrument to review. Appendix D contains a checklist which can be used for this review.

REFERENCES

Baron, J. B., & Kallick, B. (1985). What are we looking for and how can we find it? In A. L. Costa (Ed.), **Developing minds: A resource for teaching thinking** (pp. 281-287). Alexandria, VA: Association for Supervision & Curriculum Development.

Beyer, B. K. (1985, April). Critical thinking: What is it? **Social Education**, 270-276.

Brennan, M., & Stenzel, N. (1985, Spring). **ASAP Notes**. Association of State Assessment Programs Newsletter, pp. 3-4.

Costa, A. L. (1983, October 28). **Thinking: How do we know students are getting better at it?** Unpublished paper, California State University, Sacramento.

Costa, A. L. (1985). Toward a model of human intellectual functioning. In A. L. Costa (Ed.), **Developing minds: A resource for teaching thinking** (pp. 62-65). Alexandria, VI: Association for Supervision & Curriculum Development.

Costa, A. L. (Ed.). (1985). **Developing minds: A resource for teaching thinking**. Alexandria, VI: Association for Supervision & Curriculum Development.

Cox, D. W. (1985). The Purdue problem solving inventory (PEPSI), grade level and socioeconomic status: A preliminary study. **Gifted Child Quarterly**, 29, 72-73.

DeBono, E. (1977). **Teaching thinking**. Philadelphia: Transatlantic Press.

[illegible], T., & Mayhew, L. (Directors). (1954). **General education: Exploration in evaluation**. Final report of the Cooperative Study of Evaluation in General Education. Washington, DC: American Council on Education.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron, and R. J. Sternberg. **Teaching thinking skills: Theory and practice** (pp. 9-26). New York: W. H. Freeman..

Ennis, R. H., & Paulus, D. (1965). **Deductive reasoning in adolescence: Critical thinking readiness in grades 1-12, phase 1**. (ERIC ED 003818)

Frederiksen, N. (1984). Implications of cognitive theory for instruction in problem solving. **Review of Educational Research**, 54, 363-407.

Glatthorn, A. A., & Baron, J. B. (1985). The good thinker. In A. L. Costa (Ed.), **Developing minds: A resource for teaching thinking** (pp. 49-53). Alexandria, VI: Association for Supervision & Curriculum Development.

Gough, H. G. (1985). Free response measures and their relationship to scientific creativity. **Journal of Creative Behavior**, 19, 229-240.

Kearney, C. P., Kean, M. H., Roeber, E. D., Stevens, B. L., Baron, J. B., Fremer, J., & Daniel, M. (1985). **Assessing higher order thinking skills** (ERIC/TME Report 90). Princeton, NJ: ERIC Clearing House on Tests, Measurement and Evaluation, ETS.

Kneedler, P. E. (1985). **Assessment of the critical thinking skills in history/social science**. Sacramento: California State Dept. of Education.

Lipman, M. (1985). Thinking skills fostered by philosophy for children. In J. W. Segal, S. F. Chipman, & R. Glaser (Eds.), **Thinking and living skills: Vol. 1. Relating instruction to research** (pp. 83-108). Hillsdale, NJ: Lawrence Erlbaum Associates.

Marzano, R. J. (1986, September). Practicing theory. **Newsletter of the ASCD Network on Teaching Thinking**, pp. 2-3.

McPeck, J. (1981). **Critical thinking and education**. New York: St. Martins Press.

Mitchell, J. V., Jr. (Ed.). (1985). **The ninth mental measurements yearbook**. Lincoln, NE: University of Nebraska Press.

Modjeski, R. B., & Michael, W. B. (1983). **Educational and Psychological Measurement**, 43, 1187-1197.

Morante, E. A., & Ulesky, A. (1984, September). Assessment of reasoning skills. **Educational Leadership**, 71-73.

Moss, P. A., & Petrosky, A. R. (1983). **A proposal for measuring critical thinking** (ERIC ED 237399). Pittsburgh Board of Public Education.

Norris, S. P. (1985). Synthesis of research on critical thinking. **Educational Leadership**, 42:8, 40-45.

Norris, S. P., & King, R. (1984). **The design of a critical thinking test on appraising observations**. St. John's, Newfoundland: Institute for Educational Research & Development, Memorial University of Newfoundland.

Passow, A. H. Intellectual development of the gifted. In F. R. Link (Ed.), **Essays on the intellect** (pp. 23-43). Alexandria, VI: Association for Supervision and Curriculum Development.

Paul, R. (1986, August 3-6). **Dialogical thinking: Critical thought essential to the acquiring of rational knowledge & passions**. Paper presented at the Fourth International Conference on Critical Thinking & Education Reform. Washington, DC.

Perkins, D. N. (1985). What creative thinking is. In A. L. Costa (Ed.), **Developing minds: A resource for teaching thinking** (pp. 58-61). Alexandria, VI: Association for Supervision & Curriculum Development.

Presseisen, B. Z. (1985). Thinking skills: Meanings, models. materials. In A. L. Costa (Ed.), **Developing minds: A resource for teaching thinking** (pp. 43-48). Alexandria, VA: Association for Supervision & Curriculum Development.

Presseisen, B. Z. (1986, April). **Critical thinking and thinking skills: State of the art definitions and practice in public schools**. Philadelphia, PA: Research for Better Schools.

Quellmalz, E. S. (1985). Needed: Better methods for testing higher order thinking skills. **Educational Leadership**, 43, 29-35.

Reidman, H. C. (1985). Testing beyond minimums. **ASAP Notes**, Association of Assessment Programs.

Sternberg, R. J. (1984a, January). What should intelligence tests test? Implications of a triarchic theory of intelligence for intelligence testing. **Educational Researcher**, 5-15.

Sternberg, R. J. (1984b, September). How can we teach intelligence? **Educational Leadership**, 38-48.

Sternberg, R. J. (1985). Critical thinking: Its nature, measurement, and improvement. In F. R. Link (Ed.), **Essays on the intellect.** Alexandria, VA: Association for Supervision and Curriculum Development.

Sternberg, R. J. (1986). **Intelligence applied**. New York: Harcourt Brace Jovanovich.

Stewart, B. L. (1979). **Testing for critical thinking: A review of the resources**. Urbana: Bureau of Educational Research, University of Illinois. (ERIC No. ED 183588)

Stiggins, R. J., Rubel, E., & Quellmalz, E. (1986, January). **Measuring thinking skills in the classroom: A teacher's guide**. Portland, OR: Northwest Regional Educational Laboratory.

Thinking Skills Academy. (1985). Workshop materials from Research for Better Schools. Philadelphia, PA: Author.

Walsh, D. (1985, Summer). Socrates in the classroom. American Educator, 20-25.

APPENDIX A

REVIEWS OF HIGHER ORDER THINKING SKILLS TESTS

CRITICAL THINKING, PROBLEM SOLVING AND DECISION MAKING TESTS

Title of Instrument: Cornell Class Reasoning Test, Form X (1964)

Authors: Robert H. Ennis, William L. Gardner, John Guzzetta, Richard Morrow, Dieter Paulus, Lucille Ringel

Description: The authors' purpose is to test the understanding and use of eight principles of class logic in grades 4-12. The test was originally developed as part of a study on the development of formal logic. The test is not specific to a subject matter domain, but it assesses only one aspect of critical thinking--class logic. This is a multiple-choice test having one form and one level of 72 questions. The items are very structured, formal logic and require no outside information.

Authors' Description of Subtests: The test measures eight aspects of class logic.

- Whatever is a member of a class is not a non-member of that class and vice versa.
- Whatever is a member of a class is also a member of a class in which the first is included.
- Whatever is a member of a class is not (as a result of that relationship) necessarily a member of a class included in that class.
- Class exclusion is symmetric.
- Whatever is a member of a class is not a member of a class excluded from the first.
- Whatever is not a member of a class is not (as a result of that relationship) necessarily also not a member of a class in which the first is included.
- Whatever is not a member of a class is not (as a result of that relationship) necessarily a member of (nor a non-member of) another class which is excluded from the first.
- Whatever is not a member of a class is also not a member of any class included in the first.

Reliability: Test-retest reliability (based on 1964 data) for the total score is .83. This is acceptable.

Validity: The test is based on eight principles of formal logic. Correlations are moderate with ability tests and around zero for gender and SES.

Usability: The test is untimed but takes about 40 minutes to give. The test must be scored by hand. Scores are available for subcomponents as well as total score (However, no reliabilities are reported for subcomponent scores). There is no answer sheet. The only manual is ERIC No. ED 003818. The test is available packaged separately and is professionally formatted. Item difficulties are provided in grades 4, 6, 8, 10, and 12. There is no other help with interpretation. No training is required to give or score the test. The test was originally developed for use in research.

Supplemental Materials: None. Technical information is provided in Ennis, R.H. and Paulus, D. (1965). **Deductive reasoning in adolescence: Critical thinking readiness in grades 1-12, phase 1,** (ERIC No. ED 003818).

Availability: Illinois Critical Thinking Project, University of Illinois-Urbana, Champaign, IL 61820.

Comments: Stewart (1979) reviewed the instrument and found it to be a "reasonably valid and reliable measure of eight principles of class logic." He also notes that the instrument was originally used for assessing mastery--five out of six in a component area denoted mastery. This is a very structured formal logic test. The items are self-contained and require no outside knowledge. The test looks good for what it does, however most definitions of HOTS go beyond formal logic.

Title of Instrument: Cornell Conditional Reasoning Test, Form X (1964)

Author(s): Robert H. Ennis, William L. Gardiner, John Guzzetta, Richard Morrow, Dieter Paulus, and Lucille Ringel

Description: The authors' purpose is to test conditional logic for students in grades 4-12. It was originally developed as part of a study on the development of formal logic. The test is not specific to any subject matter domain. It assesses only one aspect of critical thinking--formal conditional logic. There is one form and one level of the test which has 72 items. The items are very structured, formal logic and require no outside information.

Authors' Description of Subtests: The test covers 12 subcomponents of conditional logic. Subscores are available for the subcomponents. The subcomponents are described by the author as:

- Given an if-then sentence, the affirmation of the if-part implies the affirmation of the then-part.
- Given an if-then sentence, the denial of the if-part does not by itself (as a result of its being an if-part) imply the denial of the then-part.
- Given an if-then sentence, the affirmation of the then-part does not by itself imply the affirmation of the if-part.
- Given an if-then sentence, the denial of the then-part implies the denial of the if-part.
- The if-then relationship is transitive.
- An if-then sentence implies its contrapositive.
- The if-then relation is non-symmetric.
- Given an only-if sentence, the denial of the only-if part implies the denial of the major part.
- Given an only-if sentence, the affirmation of the major part implies the affirmation of the only-if part.
- The denial or affirmation of one part of an if-and-only-if statement implies the denial or affirmation of the other part.
- Given an only-if sentence, the affirmation of the only-if part does not by itself (as a result of its being an only-if part) imply the affirmation of the major part.
- Given an only-if sentence, the denial of the major part does not by itself (as a result of its being the major part) imply the denial of the only-if part.

Reliability: The test-retest reliability of the total score is .75. This is rather low to make judgments about individuals. It is also based on 1964 data.

Validity: The test is based on 12 principles of formal logic. Correlations are moderate with ability tests and correlation with gender or SES is about zero.

Usability: The test is untimed but takes about 40 minutes to give. The test must be scored by hand. There is no answer sheet, and no manual except ERIC No. ED 003818. The test is separate but is professionally formatted. Item difficulties for about 150 students in each of grades 5, 7, 9 and 11 are provided. No other help in interpretation is given. No training is required to give or score the test. The test was originally developed for use in research.

Supplemental Materials: None. Technical information is provided in Ennis, R. H., & Paulus, D. (1965). **Deductive reasoning in adolescence: Critical thinking readiness in grades 1-12**. (ERIC No. ED 003818).

Availability: IL Critical Thinking Project, University of Illinois-Urbana, Champaign, IL 61820.

Comments: Stewart (1979) reviewed the test and found it to be a "reasonably valid and reliable measure of mastery of 12 principles of conditional logic." He notes that the instrument was originally used by the author to assess mastery--a score of 5 out of 6 on each component denotes mastery. This is a very structured formal logic test. The items are self-contained and require no outside knowledge. The test is good for what it does, however most definitions of HOTS go beyond formal logic.

Title of Instrument: Cornell Critical Thinking Tests, Level X and Level Z, third edition (1985)

Authors: Robert H. Ennis and Jason Millman

Description: The authors' purpose is to test general critical thinking skills. Tests were designed to assist in conceptualizing critical thinking and for use in the schools. Critical thinking is defined as "the process of reasonably deciding what to do." Form X is designed for grades 4-12; form Z is designed for adults. These tests are not specific to a subject matter domain and are intended to cover critical thinking skills in general. There is one form of each multiple-choice test. Level X has 71 items and Level Z has 52 items. Some of these are self-contained formal logic items and others are intended to relate to situations in everyday life.

Authors' Description of Subtests: The authors list the aspects of critical thinking included, but do not define them:

- Induction
- Deduction
- Value Judgment
- Observation
- Credibility
- Assumption

Reliability: Internal consistency reliabilities (split half and KR-20) range from .67 to .90 for Level X (median = 80) based on 3500 students, and ranged from .50 to .77 for Level Z (median = .67) based on 2000 adults. The author recommends against profiling individual students on subtest scores because of their short length and consequent low reliability.

Validity: Content is based on Ennis' (1987) conception of critical thinking. Items were reviewed for correct keyed response. Correlations with other critical thinking tests, ability tests and achievement tests are about .5. Correlation with gender, SES, and other affective measures are about zero. The authors present some studies showing the relationship of training programs to changes in scores. Results of factor analysis studies are inconclusive. The authors conclude "there is no definitive establishment of the construct validity of Level Z or of any critical thinking test for that matter." Regardless of this gloomy self-description, the Cornell seems to be one of the better measures in terms of examination and discussion of issues regarding the use of structured format tests in assessing higher order thinking skills. Four reviews of the instrument (Ennis, 1986; Modeski and Michael, 1983; Stewart, 1979; McPeck, 1981) agree that the biggest issue is probably that cases could logically be made to support other answers.

Usability: The tests require about 50 minutes to give. The tests can be machine or hand-scored. There is a correction for guessing. The materials are professionally packaged. No training is required to give or score the tests. Means and quartiles for individual groups of students tested between 1960 and 1980 are provided. These are, however, based on small numbers of students. Norms are not comprehensive. It was developed for use by schools.

Availability: Midwest Publications, P.O. Box 448, Pacific Grove, CA 93950.

Comments: There is a good, clear discussion of some assessment issues and how the Cornell attempts to deal (or not deal) with them. Fairly frank and self-revealing. One of the better instruments because of this feature.

Title of Instrument: Ennis-Weir Critical Thinking Essay Test (1985)

Authors: Robert H. Ennis and Eric Weir

Description: The authors' purpose is to test students' ability to analyze logical weaknesses in arguments by responding to a fictional letter. The test is recommended for use in grades 9 through adult. This test is not specific to any subject matter domain and is intended to be a general measure of critical thinking. This is an essay test having one form and one level. The student responds in writing to eight paragraphs, each of which has a flaw in reasoning.

Authors' Description of Subtests: Although the test really has no subtests, areas of critical thinking competence covered by the Ennis-Weir are:

- Getting to the point
- Seeing the reasons and assumptions
- Stating one's point
- Offering good reasons
- Seeing other possibilities (including other possible explanations)

Responding appropriately to and/or avoiding:

- Equivocation
- Irrelevance
- Circularity--Reversal of an if-then (or other conditional relationship)
- The straw person fallacy
- Overgeneralization
- Excessive skepticism
- Credibility problems
- The use of emotive language to persuade

Reliability: Since this is an open-response test, the authors report interrater reliabilities. Two samples of size 27 and 28 have interrater reliabilities of .86 and .82. These are reasonable, but the samples are small and non-representative.

Validity: There is only discussion of content validity. The test is based on critical thinking competencies in Ennis' taxonomy (1987). The author feels that predictions and concurrent validity are not possible because there exists no established outside criterion for the ability the test was designed to measure.

Usability: The test requires about 1 hour and 10 minutes to give. No training is required to give the test, but it must be hand scored by trained scorers. The manual provides detailed statements about what could be included in responses. Packaging is attractive. The manual has means for 55 college and 8th grade students. A little help is given with interpreting scores, but there are no guidelines on standards. The test was developed for use by schools and for research.

Supplemental Materials: A manual includes guidance for scoring each paragraph of the essay.

Availability: Midwest Publications, P.O. Box 448, Pacific Grove, CA 93950.

Comments: This test represents an attempt to get around problems with multiple choice tests. It attempts to present a real-world fuzzy problem that is familiar to most students--parking. It might not work internationally. The only review found was by Stephen Norris at a recent conference in critical thinking. He stated that the guide to scoring the Ennis-Weir test stresses conclusions rather than reasoning and so falls into the same trap as multiple-choice tests. This reviewer does not agree with that assessment. Although the test represents a good attempt, there is not much in terms of validity, standards of comparison or help in interpreting/using results.

Title of Instrument: Judgment: Deductive Logic and Assumption Recognition, Grades 7-12 (1971)

Authors: Edith Shaffer and JoAnn Steiger

Description: The authors' purpose is to assess the logical ability of students. The rationale is that if a student cannot correctly interpret logical problems when given full data, his ability to deal with more difficult situations . . . is probably limited . . ." There are five multiple choice tests in the booklet, each of which measures a separate aspect of critical thinking. None of the tests is specific to a subject matter domain. The tests are intended for use in grades 7-12.

Authors' Description of Subtests:

- Conditional Reasoning Index: This measure deals with a particular aspect of formal logic: "if-then" statements (part of deduction). Some of the items deal with subjects which may be emotionally laden for the student.

- Class Reasoning Index: This measure also deals with one element of formal logic: "all, none, and some" statements (part of deduction). Separate scoring may also be done for items with emotionally-laden content.

- [illegible] This measure consists of brief statements each followed by a list of proposed assumptions from which the student must choose the appropriate ones.

- Assumption Recognition Index II: Here the student must read a several-sentence argument (perhaps emotionally charged for him) and then select the appropriate assumptions from a list of suggested ones.

- Recognizing Reliable Observations: Deals with the ability to weigh evidence by evaluating the source.

Reliability: No information given.

Validity: The tests were based on general principals of formal logic as outlined by various authors, especially Ennis (1965). Each test was reviewed by two content experts. There is a general statement on quality control: "collections that contain complete measures are field-tested for purposes of development prior to publication," (p. vii). However no other information or results of this process is presented.

Usability: The time required to give each test is: Conditional Reasoning Index--40 minutes; Recognizing Reliable Observations--15 minutes; Class Reasoning Index--35 minutes; Assumption Recognition Index I--15 minutes; Assumption Recognition Index II--20 minutes. The user is responsible for setting up a scoring system. Tests are bound into a 6" x 9" booklet. Use would probably require recopying. There are no norms and no help with interpreting results. No training is required to give or score the test. The tests were developed for use by schools.

Availability: IOX Assessment Associates, Box 24095, Los Angeles, CA 90024.

Comments: The tests are recommended by the authors for group assessment, not individual student diagnosis. Users are urged by the authors not to infer general judgment ability from these few aspect specific tests. On the logic tests, half of the items are emotionally laden and half are not.

There is little evidence of validation. Stewart (1979) reviewed these instruments and felt that they generally needed further development (They seem to have not been further developed since then). He questioned whether the items measured the skill intended, and if some items may have more than one right answer.

Title of Instrument: Means-Ends Problem Solving (1975)

Authors: Jerome J. Platt and George Spivack

Description: The authors' purpose is to measure the individual's ability to orient himself to and conceptualize means of moving towards a goal, specifically in the area of problem solving in interpersonal relationships. The instrument is intended for use in grades 9 to adult. There is a children's form for grades 5-7. The test is not specific to any subject matter domain. It is meant as a general measure of interpersonal problem solving. This is an open-ended interview or essay in which examinees are presented with 10 situations in which an interpersonal conflict exists. The prompts provide a situation and a solution. The examinee must outline how the protagonist could have moved from the original situation to the solution.

Authors' Description of Subtests: There are no subtests, but the protocol is scored for individual steps in problem solving (means), awareness of potential obstacles, and awareness of the passage of time.

Reliability: Interrater reliabilities for nine stories and 15 students was .98. Test-retest reliability ranged from .43-.64 (2 1/2 weeks to 8 months) in 3 samples (total N = 73). Internal consistency reliability between stories ranged from .80 to .84. These reliabilities look pretty good.

Validity: The instrument was developed to fill a gap in work on problem solving--problem solving skills for interpersonal situations. The instrument differentiates between normal individuals and those needing psychiatric help, and between those with various levels of social competence. There are small to moderate correlations with intelligence test scores. Factor analyses suggest that the stories measure the same quality of thinking. Several groups of normal adults agreed on what were effective strategies for moving from the problem to the solution.

Usability: The **Means-Ends** procedure is untimed and no time estimates are given. Because the method is open-ended, scorers must be carefully trained. There is a long section in the manual devoted to scoring. The instrument and manual are not commercially packaged--materials must be copied from the manual. For comparison purposes, mean scores for the various prompts are given for 6 male and 6 female groups (students, hospital employees and psychiatric patients) ranging in size from 23 to 54. The test was originally developed for use in research and for use with maladjusted adolescents.

Supplemental Materials: A manual includes a description of the instruments, administration and scoring, summary of research, and scoring sheets.

Availability: Department of Psychiatry, School of Osteopathic Medicine, University of Medicine and Dentistry of New Jersey, 401 Haddon Avenue, Camden, NJ 08103.

Comments: This instrument was originally developed for use with psychiatric patients to determine their interpersonal problem solving skills. It has since been used with a number of different groups. Even though it is clinically oriented, we included it because of its view of problem solving in the interpersonal domain. The instrument has been used extensively in research and has a great deal of evidence accumulated that it distinguishes between groups and predicts behavior. However, the groups it distinguishes between are very disparate (e.g., normal adolescents and patients in a psychiatric ward). It may not be as useful in distinguishing between groups that are in the schools (i.e., all normal).

Title of Instrument: New Jersey Test of Reasoning Skills, Form B (1985)

Author: Virginia Shipman

Description: The author's purpose is to test general reasoning ability with a low-reading level instrument. The test is intended for students in grades 4-College. The test is not specific to any subject matter domain and is intended to assess many aspects of reasoning. This multiple-choice test has one form and one level of 50 questions.

Authors' Description of Subtests: The test is intended to measure 22 reasoning skills:

- Converting statements
- Translating into logical form
- Inclusion/exclusion
- Recognizing improper questions
- Avoiding jumping to conclusions
- Analogical reasoning
- Detecting underlying assumptions
- Eliminating alternatives
- Inductive reasoning
- Reasoning with relationships
- Detecting ambiguities
- Discerning causal relationships
- Identifying good reasons
- Recognizing symmetrical relationships
- Syllogistic reasoning (categorical)
- Distinguishing differences of kind and degree
- Recognizing transitive relationships
- Recognizing dubious authority
- Reasoning with 4-possibilities matrix
- Contradicting statements
- Whole-part and part-whole reasoning
- Syllogistic reasoning (conditional)

Reliability: Based on a subsample of 2,346 students in a pilot sample, the internal consistency reliability of the total score is .84-.94. (This information could be from an earlier experimental version of the test.) The reliability is quite good. No reliabilities are provided for individual skills.

Validity: The author developed a taxonomy of logical operations performed in childhood based on a survey of logical competencies produced by language acquisition. She selected 22 of these (deduction and induction mostly) for the test. The author did not want this to be a test of reading comprehension so she kept the reading level at grade 5 or below. Correlation with subject matter tests are fairly high, especially reading tests. (Usually test developers do not want correlations with subject areas to be this high because then the test looks like an achievement test.) The author offer no interpretation of these correlations. She claims there are no items which depend on recall of content or information outside the problem itself. However, we identified many items that could be affected by the general knowledge of the test taker. Also some were found which could be answered from general knowledge and not the logic involved, and some where the test taker might be confused whether to use general knowledge or only information in the item. There are also at least two dry vocabulary items.

The author presents no evidence that these confusions do not occur when the test is used and does not discuss these as potential problems. The author also provides no evidence that the test items measure what they claim to measure.

Usability: The test is untimed and can be given in 30-60 minutes. The test can only be machine-scored by the publisher. The tests look nice. The only help on interpretation is student score means based on the students tested to date. No training is required to give the test. It was developed for use by schools and in research. The tests are "rented" to users. Test booklets must be returned to the publisher within 12 months. A single price per booklet covers rental and scoring. The test is intended to accompany the Philosophy for Children program.

Supplemental Materials: Background Paper (1983) and information on the Philosophy for Children program as well as information about the test. There is no scoring key, administration manual, or information on how to interpret results.

Availability: Institute for the Advancement of Philosophy for Children, Montclair State College, Upper Montclair, NJ 07043.

Comments: The test provides subscores on individual components even with few items per component and no estimates of the reliabilities of these scores. There are alternate forms being prepared. There is no review of this test in Mental Measurements Yearbook.

Title of Instrument: Purdue Elementary Problem Solving Inventory (1972)

Authors: John Feldhusen and John Houtz

Description of the Test: The authors' purpose is to test grade 2 to 6 students' ability to solve commonsense real-life problems. This test is not specific to any subject matter domain and is meant as a general measure of problem solving. There is one form and one level containing 49 multiple-choice questions.

Authors' Description of Subtests: students are shown a cartoon of a situation and are tested for:

- Sensing the problem: if there is or is not a problem.
- Identifying a problem: one statement which specifies the problem.
- Asking questions: pick from each set of three questions the question which would be most useful in clarifying the problem.
- Guessing causes: pick from a set of three possible causes the one which would most likely be the cause of the problem.
- Clarification of goal: given an ambiguous goal or task, select the piece of information which would clarify the goal or an adequate search model.
- Judging if more information is needed: whether sufficient information is or is not available to proceed to a solution.
- Analyzing details of the problem and identifying critical elements.
- Redefinition or transformation of common objects in order to see their potential use.
- Seeing implications: pick the most likely result if the given solution were implemented.
- Verification: pick an appropriate method.
- Solving a single solution problem: pick the alternative which will solve the problem.
- Solving a multiple solution problem: picking unusual and best solutions to a problem with multiple steps.

Reliability: Based on 1,073 students in Indiana, the internal consistency reliability of the total score is .79.

Validity: The authors reviewed the problem solving literature for content and format. There is some information to show that the format is appropriate for the age group. A factor analysis showed one main problem solving factor. No other information is given.

Usability: The test takes about 40-45 minutes to give. All questions and response choices are read to students using a tape. The item situations are shown to students on a filmstrip. It is not professionally packaged. Norms are available by sex in grades 2, 4, and 6 (N=571). It was developed for use in the schools.

Supplemental materials included: Research articles between 1972 and 1985 using the test.

Availability: Gifted Education Resource Institute, Purdue University, South Campus Courts, Building G, West Lafayette, IN 47907.

Comments: Some items are ambiguous. Answers would depend on the level of sophistication of examinee or knowledge/experience. Others seem to depend on ability to notice details in the pictures, memory and language ability. However, there is some evidence that the **Purdue** assesses logical thinking and concept formation. One review (Cox, 1985) judged the **Purdue** suitable for grades 2-6 of all SES levels. They felt it had potential utility for problem solving programs.

Title of Instrument: Ross Test of Higher Cognitive Processes (1976)

Authors: Catherine M. Ross and John D. Ross

Description: The authors' purpose is to assess the higher level thinking skills of students in the intermediate grades. The test is intended for use in grades 4-6. The test is not specific to any subject matter and is intended to measure several aspects of higher level thinking. There is one form and one level containing 105 items. Items are mostly multiple-choice.

Authors' Description of Subtests: The authors relate each subtest to Bloom's Taxonomy of Educational Objectives, Handbook I. The part of the taxonomy is in quotes below.

- Analogies: This section consists of 14 items which measure a student's ability to perceive analogous relationships between pairs of words. It relates to "Analysis of Relationships,"
- Deductive Reasoning: This section consists of 18 items which measure a student's ability to analyze statements in logic. It relates to "Judgments in Terms of Internal Evidence."
- Missing Premises: This section contains eight items which measure a student's ability to identify the missing premise needed to complete a logical syllogism, when given only one premise and a conclusion. It relates to "Analysis of Elements."
- Abstract Relations: This section contains 14 items which measure a student's ability to study data and synthesize a logically consistent scheme for organizing them to form a conceptual structure. It relates to "Derivation of a Set of Abstract Relations."
- Sequential Synthesis: This section measures a student's ability to organize ideas into a coherent communication. It relates to the "Production of a Unique Communication."
- Questioning Strategies: This section measures a student's ability to evaluate methods of obtaining data by judging the efficiency of the method in producing the best data. It relates to "Judgments in Terms of External Criteria."
- Analysis of Relevant and Irrelevant Information: This section measures a student's ability to analyze data and identify critical information or the lack of same. It relates to "Analysis of Relationships."
- Analysis of Attributes: This section presents groups of similar figures which have a variety of features, or attributes. Possession of a distinct combination of attributes designates a figure as a member of a set. This section relates to "Derivation of a Set of Abstract Relations."

Reliability: Based on the standardization sample (see norms below) internal consistency reliability for the total score is .92; test-retest reliability (3 days apart) is .94. Total score reliabilities are good. No reliabilities are reported for subtests.

Validity: The test was designed to measure the higher level skills in Bloom's taxonomy-analysis, synthesis and evaluation, especially those that deal with verbal abstractions. Correlation of scores with age was .64. The test distinguished between groups of gifted and non-gifted students identified by another process. Correlation with an IQ test was small (.16-.40). Items were selected based on traditional item statistics. There is no evidence presented that the item types actually measure Bloom's categories or that performance on them is related to good thinking or achievement. There is also no evidence that students understand the task, that it is not influenced by reading comprehension, or that the scales measure independent factors.

Usability: The test is designed to be given in 2 sittings of about an hour each (excluding instructions). The test can be hand or machine scored. Hand scoring using the key takes about 10 minutes per test; using a supplemental overlay takes about 5 minutes per test. A

scoring program is also available. The test and other materials are nicely packaged. The standardization sample was 527 gifted and 610 non-gifted students in 9 districts in the state of Washington. There were about 100 students per grade in each of the gifted/non-gifted samples. Because the norms are based on so few students, and because there are only 8-18 items in each subtest, the norms are not very discriminating. For example, scores of 0-3 on analogies are all a percentile of 1, while a jump from a score of 8 to 9 gives an 18 percentile point jump. Means are provided for all groups and item statistics are provided for all items. No training is required to give or score the test. The test was developed for use by schools.

Supplemental Materials: A manual contains information on test content, directions for administration and scoring, some technical information on the test, scoring keys and norms. Separate scoring overlays and answer sheets are also available.

Availability: Academic Therapy Publications, 20 Commercial Blvd., Novato, CA 94947-6191.

Comments: The authors specify that the test can be used to screen students for inclusion in a special program, assess the effectiveness of a program, or assess students' higher-level thinking skills. Many of the tasks on the test are the same as those found in formal logic (e.g., missing premises, deductive reasoning) or general intelligence tests (e.g., analogies, abstract relations, analysis of attributes). Many of the item types are abstract and not representative of real-life problem solving or critical thinking.

In general, administration and scoring is easy and the manual is easy to use. Weaknesses include lack of information on reliability and validity and lengthy administration time (Mitchell, J.V., 1985, #1061) Although the norms are incomplete, more comparative information is presented than with many other tests.

Title of Instrument: Test on Appraising Observations (1983)

Authors: Stephen P. Norris and Ruth King

Description: The authors' purpose is to assess students' ability to appraise the reliability of observational statements. The test is intended for use with junior high school students to adults, but it is best for senior high students. This multiple-choice test has one form and one level of 50 items. In each question, students decide which of two statements, if either, is more believable given the context. Examinees give direction of endorsement only instead of degree (as on the **Watson-Glaser**) to help avoid problems with level of sophistication of the test taker.

Authors' Description of Subtests: The test does not really have subtests although it is designed to cover 31 principles of deciding on the validity of an observation. Subscores can be computed on the major categories. The basic principle is that observation statements tend to be more believable than inferences based upon them. Other principles relate to characteristics of the observer, the observation condition, and the observation statement itself.

Observer. An observation statement tends to be believable to the extent that the observer:
- is functioning at a moderate level of emotional arousal;
- is alert;
- has no conflict of interest;
- is skilled at observing the sort of thing observed;
- has a theoretical understanding of the thing observed;
- has senses that function normally;
- has a reputation for being honest and correct;
- uses as precise a technique as is appropriate;
- is skilled in the technique being used;
- has no preconceived notions about the way the observation will turn out;
- was not exposed, after the event, to further information relevant to describing it; and
- is mature.

Observation Conditions. An observation statement tends to be believable to the extent that the observation conditions provide:
- a satisfactory medium of observation;
- sufficient time for observation;
- more than one opportunity to observe; and
- adequate instrumentation, if instrumentation is used.

Observation Statement. An observation statement tends to be believable to the extent that it:
- commits the speaker to holding a small number of things to be true;
- is corroborated;
- is no more precise than can be justified by the observation technique being used;
- is made close to the time of observing;
- is made by the person who did the observing;
- is strongly believed to be corroborated by the person making it;
- does not conflict with other statements for which good reasons can be given;
- is made in the same environment as the one in which the observation was made;
- is not about an emotionally-loaded event;
- is the first report of the event provided by the speaker;
- is not given in response to a leading question;

- does not report a recollection of something previously forgotten;
- reports on salient features of an event; and
- is based upon a reliable record, if it is based upon a record.

Reliability: Internal consistency reliability of the total score is .69. The authors caution against using part scores for individual students because of the low reliability.

Validity: The test is based on Robert Ennis' principles of appraising observations. The authors cite research that supports these principles. The authors developed questions by studying the mental processes of people while they responded to the questions. They tried to check understanding of the task and why people chose the answers they did. They also attempted to check for other irrelevant influences such as testwiseness, readability and clarity of directions. Correlations with other tests of critical thinking range from .08 to .74, depending upon the group sampled.

Usability: The test is untimed, but requires about one class period. The test must be hand scored. The test is professionally packaged. The manual provides means for 500 high school students in Ontario, Canada, and decile norms. There are no mastery criteria. No training is required to give or score the test. The test was developed for use by schools.

Supplemental Materials: **The Design of a Critical Thinking Test on Appraising Observations** (Norris and King, 1984) is also available. This report includes a detailed description of the test development process and protocols for interviewing students about their responses.

Availability: Institute for Educational Research and Development, Memorial University of Newfoundland, St. John's, Newfoundland, CANADA A1B 3X8.

Comments: This test attempted to circumvent some current issues when using paper and pencil tests to measure HOTS. First, it tries to present real-life, fuzzy problems. It tries to embed each situation into a story to avoid having to draw conclusions on too little data. The authors also attempted to validate the test by examining the mental processes people use to respond. This is based on the idea that if mental processes which are suitable lead to good test performance and unsuitable ones lead to poor performance, then the test is valid. This instrument has the most extensive attempt at validation of all those reviewed. The instrument looks good. But, it only is intended to measure one aspect of higher order thinking-- appraising observations. The author has alerted us to other instruments which are under development as part of the Memorial University Critical Thinking Test Series. They include: **Essay Test on Appraising Observations; Essay Test of Inductive Reasoning Strategies;** and **Test of Principles of Inductive Reasoning.**

Title of Instrument: Think It Through (1976)

Authors: Not specified

Description: The author's purpose is to measure a young child's problem solving ability. The content is not specific to any single subject matter and it is intended to measure several problem solving skills. It is a group test. Level A has 32 and level B 31 multiple-choice questions. Children mark their choices right in the test booklet. Questions are read to students.

Author's Description of the Subtests:

- Classification: Assesses the child's ability to discriminate among the features of objects by classifying them on the basis of their physical properties.
- Solution Evaluation/Time Sequence: Measures the child's ability to judge the appropriateness and consequences of several different solutions and to judge which of three pictures represent the beginning of a sequence.
- Word Problems: classification and sorting items that require the child to identify some factor several objects have in common; and items that require the child to solve problems where conventional objects must be viewed in unconventional ways.
- Patterns: Patterns to be completed either in the form of sequences of beads or of "broken" plates to be repaired.
- Mazes: Items that present three paths to a goal and the child selects the quickest route to the goal.

Reliability: Internal consistency reliability for form A total score ranged from .81 to .82. Subtests ranged from .61 to .76. Internal consistency reliability for level B total score ranged from .64 to .75. Subtest reliabilities are too low for profiling individual students. However, these reliabilities appear good for this age group.

Validity: Students with no preschool experience had lower scores than those with preschool experience. There were no differences between boys and girls. No rationale is provided for the item types selected and no evidence is reported that scores predict any kind of school performance or problem solving performance in real life.

Usability: The test is untimed since items are read to students. It takes 30-40 minutes to give. The test can be machine or hand scored. The test is part of the CIRCUS achievement test series. Packaging is professional. Means and standard deviations for ages 4.4 to 6.0 are provided for each subtest. This information is also presented by region, race, SES and preschool experience. Item statistics are provided, as well as suggested verbal interpretations of various score ranges (e.g. "very competent"). The percent of students in nursery school and kindergarten falling into each of these ranges is given. School means are provided for preK, K and 1. Expected level B performance based on level A scores is given. No training is required to give or score the test. The test was developed for use by schools.

Supplemental Materials: Scoring key, class performance record, sentence report table (for converting numerical scores into written text. Teacher rating inventory (for identifying children who appeared to have difficulty in coping with the task).

Availability: CTB/McGraw Hill, Del Monte Research Park, 25000 Garden Rd., Monterey, CA 93940.

Comments: The authors recommend that the test not be used with students younger than age 4. It may be difficult to give this as a group test in nursery school. Level B has no total score. In level A, the first 6 items are not in any subtest. The strength of this test is the variety of ways presented for interpreting scores. The weakness is the lack of information on validity.

Title of Instrument: Watson-Glaser Critical Thinking Appraisal (1980)

Authors: Goodwin Watson and Edward Glaser

Description: The authors' purpose is to measure some of the abilities involved in critical thinking. They suggest it be used in schools for student/classroom diagnosis and program evaluation, and selection of candidates for positions (both within schools and outside of schools) requiring critical thinking. It is intended for use in grades 9 to adult. It is not specific to any content area domain and is intended to measure general critical thinking ability in real-life situations. There are two forms, each of which has 80 items. The items require judgment about real-life situations.

Authors' Description of Subtests:

- Inference: discriminating among degrees of truth or falsity of inferences drawn from given data.
- Recognition of Assumptions: recognizing unstated assumptions or presuppositions in given statements or assertions.
- Deduction: determining whether certain conclusions necessarily follow from information in given statements or premises.
- Interpretation: weighing evidence and deciding if generalizations or conclusions based on the given data are warranted.
- Evaluation of Arguments: distinguishing between arguments that are strong and relevant and those that are weak or irrelevant to a particular question at issue.

Reliability: Based on 11 groups of students in high school and college (of 66 to 243 students each), internal consistency reliabilities for subtests ranged from .69 to .85; test-retest reliability was .73 (N = 96); alternate form reliability is .75 (N = 228). Reliabilities on subtests may be too low for individual student profiling.

Validity: All passages are at a 9th grade readability or below as measured by the Chall, Fry, and Flesch formulas. Content is based on Dressel and Mayhew's (1954) conception of behaviors related to critical thinking. "Judgments of qualified persons and results of research studies . . . support the belief that the items in the critical thinking abilities represent an adequate sample of (those) five abilities." The authors report several studies that scores increase after participating in educational programs in which critical thinking was emphasized. Correlations with aptitude scores range from .29 to .81 and with achievement from .12 to .50. (There are higher correlations with verbal than with computational scores.) The authors present information on factor analyses which show that the items clustered on one factor which is different than general intelligence. There is no evidence on how level of sophistication or philosophical differences affect scores. This is especially important since some items require "common knowledge". The authors expect that a United States citizen with a ninth grade reading ability and an ability to think critically should understand the situations presented in the items. The directions are complex in that they require the examinee to understand logical distinctions (such as truth beyond a reasonable doubt) which may also be affected by level of sophistication (McPeck, 1981, Modjeski and Michael, 1983, Stewart, 1979). Dr. Glaser would disagree with this evaluation. The test is a test of sophisticated thinking so any attempt to show a correlation between a student's score and level of sophistication is redundant. The questions were tested on leaders in education, psychology, philosophy and business and fine-tuned until all of the sample group agreed with the key.

Usability: The test takes 40 minutes to give. The test can be hand scored or machine scored by the publisher. It is professionally packaged. Norms are provided for grades 9-12

based on samples of 1700-2000 students per grade. These seem adequate. There are also norms for other special groups (but with such smaller Ns) - college students and various professional groups. No training is required to give or score the test. It was developed for use by schools and other practical settings.

Supplemental Materials: Scoring template; examinee record form for summarizing scores across a classroom; separate answer sheet; and a manual containing a description of the test, test administration, scoring, interpretation and development information.

Availability: Psychological Corporation, 555 Academic Court, San Antonio, TX 78204-0952.

Comments: Other forms are Ym and Zm. These are the old forms of the test and require about 50 minutes to give. Half the situations in the current forms are noncontroversial and half are controversial. Instructions may be confusing to test takers. This is one of the oldest and most used tests of critical thinking.

OTHER CRITICAL THINKING, PROBLEM SOLVING OR DECISION MAKING TESTS

Applications of Generalizations Test (1969) by Norman E. Wallen. Available from: Tests in Microfiche (#008426) ETS Test Collection, Princeton, NJ 08541-0001. Grades 4-12. The Taba curriculum project developed this test to assess students' ability to use generalizations after participating in the Taba social studies curriculum where they learned several widely used generalizations about the history of civilization. There are 65 questions for a single level. Students indicate whether a statement is "probably true" or "probably false." There are no norms and scoring interpretation is tied to the curriculum itself.

Primary Test of Higher Processes Thinking (1978) by Winnie V. Williams. Available from Tests in Microfiche (#013161) ETS Test Collection, Princeton, NJ 08541-0001. The author developed this test to determine cognitive abilities in the higher levels of thinking. It is intended for use in grades 2-4 and was originally used with gifted students. It is a general knowledge test and includes the following subtests: convergent production and analogies, sequential relationships, logic, deductive reasoning and divergent thinking. There is no rationale given for the item types selected or the scoring system used, and no information on validity.

TAB Science Test: An Inventory of Science Methods (1966), by David P. Butts, University of Texas. Available from Tests in Microfiche (#007741) ETS Test Collection, Princeton, NJ 08541-0001. The author developed a unique observational procedure for testing the order in which a student solves a problem, using "tabs" to keep track of each student's progress. The test is appropriate for grades 4-6. It tests searching, data processing, discovery, verification and application.

Test of Enquiry Skills (1979) by Barry J. Fraser. Available from Australian Council for Educational Research, Frederick St., Hawthorn, Victoria 3122, Australia. Grades 7-10. This test is long and only one section refers to higher order thinking skills. There are 87 questions in science, social studies and general studies measuring nine skills which are grouped under the headings: "Using Reference Materials", "Interpreting and Processing Information" and "Critical Thinking in Science". It is designed for grades 7-10. Reliabilities of subtests range from .57 to .83.

Test of Science Comprehension (1963), by Clarence H. Nelson and John M. Mason, Michigan State University. Available in: A Test of Science Comprehension for Upper Elementary Grades. (1963) **Science Education**, 47, p319-330. Critical thinking questions ask students to look at graphic material from the sciences and interpret data and draw conclusions. It is written for grades 4-6 but seems appropriate for higher grades as well.

DEVELOPMENTAL TESTS

Title of Instrument: Arlin Test of Formal Reasoning (ATFR-1984)

Author: Patricia Kennedy Arlin, Ph.D.

Description: The author's purpose is to provide a quick way to assess students' level of cognitive development according to Piaget's stages of formal operations. It is intended for grades 6 through adult. This instrument is not specific to a subject matter domain. It assesses only one aspect of HOTS: Piaget's stage of formal operations. It is a multiple-choice test with 32 questions on one form and one level. These are problem-solving questions using math and science concepts applied to everyday life. Some items have follow-up questions which ask why the student choose the answer he or she did. This is an attempt to get at the process of problem solving in addition to coming up with a right answer.

Authors' Description of Subtests:

- Multiplicative Compensations: understanding that when there are two or more dimensions to be considered in a problem, gains or losses in one dimension are made up for by gains or losses in the other dimensions.
- Probability: the ability to develop a relationship between the confirming and the possible cases.
- Correlations: the ability of a student to conclude that there is or is not a causal relationship, whether negative or positive, and to explain the minority cases by inference of chance variables.
- Combinational Reasoning: the concept of generating all possible combinations of a given number of variables, choices, events, or scenarios when a problem's solution requires that all possibilities be accounted for.
- Proportional Reasoning: a mathematical concept which involves the ability to discover the equality of two ratios which form a proportion.
- Forms of Conservation Beyond Direct Verification: the ability to deduce and verify certain conservations by observing their effects and thus inferring their existence.
- Mechanical Equilibrium: the ability to simultaneously make the distinction between and the coordination of two complementary forms of reversibility--reciprocity and inversion.
- The Coordination of Two or More Systems or Frames of Reference: the concept which requires the ability to coordinate two systems, each involving a direct and an inverse operation, but with one of the systems in a relation of compensation or symmetry in terms of the other. It represents a type of relativity of thought.

Reliability: The authors tested 7,212 students in 6 states. Internal consistency reliabilities for the total score ranged from .60 to .73. Reliabilities on subcomponents may be too low for profiling individual students.

Validity: Test content is based on Piaget's theories of development. Development of the test included keeping only those items which produced "results comparable to an individual's performance on the Piagetian clinical tasks." Other validity studies were done on an earlier form of the test. A review (#80) in **Mental Measurements Yearbook** (Mitchell, J.V., 1985) of this test concluded that "the total score assessment provided by the ATFR is reasonably well correlated with level of formal operational functioning." The reviewer, however, did quibble with some of the test's definitions of levels of formal operations, and felt that it is

theoretically meaningless to assess subcomponents because formal operations is a holistic concept. This is a debatable issue (see Inhelder and Piaget (1958) **The Growth of Logical Reasoning from Childhood to Adolescence**, p308).

Usability: The test is untimed, but usually takes about 45 minutes. Scores are given for overall level of formal reasoning and subscores for the eight components. The test can be machine or hand-scored. Eight different templates are required to score the eight subcomponents since items are not together. This would make the test somewhat awkward to hand score. The test is attractively packaged. Interpretation is criterion-referenced--tied to Piaget's levels of formal operations. However, average test performance for grades 6-12 is provided (based on a large sample).

Supplemental Materials: Scoring templates for total and each subtest; manual; computer reporting with Apple or IBM computers; workbook series for applying the ATFR in the classroom.

Availability: Slosson Education Publications, P.O. Box 280, East Aurora, NY 14052.

Title of Instrument: Understanding in Science (1975)

Authors: R.P. Tisher and L.G. Dale

Description: The authors' purpose is to provide a paper and pencil alternative to standard Piagetian clinical interviews to measure concrete and formal operational thinking. It is recommended for use in grades 7-9. The situations presented relate to basic science concepts. There is one form and one level having 24 questions. Most questions are multiple choice, others require marking a diagram or writing a short response.

Authors' Description of Subtests: There are no subtest scores, but the situations tested are reflection from a plane, balance, balancing columns of liquids, and projection of shadows.

Reliability: None provided.

Validity: The test is based on four experimental situations described by Inhelder and Piaget. Fifty-seven grade 7-9 students were given the test and were clinically interviewed using procedures described by Inhelder and Piaget. There was a 77% agreement in the level of operational thinking displayed by the students.

Usability: The test takes about 40 minutes to give. The test must be hand scored. The materials are packaged for reproduction by the users. There are no norms. Guidance for interpreting results includes labeling a score by the Piagetian stage it represents--early concrete, late concrete, early formal or late formal stage. These levels are tied to instructional materials developed by the Australian Science Education Project. The test was originally developed for use in research.

Supplemental Materials: A short manual defines terms and gives administration and scoring instruction. A separate answer sheet is available.

Availability: Australian Council for Education Research, Ltd., Frederick Street, Hawthorn, Victoria 3122, Australia.

Comments: The authors specify that this is an experimental instrument and users should be extremely cautious when using the instrument. They have made it available because of its potential usefulness. The authors have described it further in a chapter in the **Third Handbook of Research on Teaching.**

OTHER DEVELOPMENTAL TESTS

Classroom Test of Formal Reasoning (1978) by Anton E. Lawson, University of California-Berkeley. Available from The Development and Validation of a Classroom Test of Formal Reasoning, **Journal of Research in Science Teaching**, 5, 11-24. Fifteen items are described for testing Piaget's formal operational thinking in grades 8-12. Questions test: combinations, correlations, probability and proportions. Students are classified into Piaget's developmental levels according to test scores.

Formal Operations Measure (n.d.), Carol Ann Tomlinson-Keasey, University of California-Riverside. Available from Tests in Microfiche (#010271) ETS Test Collection, Princeton, NJ 08541-0001. Two forms are available for pre- and post-testing this Piagetian instrument which tests formal operational thinking or abstract reasoning for college students. The seven tasks are designed as experiments testing proportionality, systematic searches, isolation of variables, analogies, correlations, abstractions and probability. The questions are open-ended and ask for a student's reasoning. No information on reliability or validity is provided.

Formal Operations Test (Biology, History and Literature) (1979) William M. Bart, University of Minnesota. Available from Tests in Microfiche (008422, 008423, 008424) ETS Test Collection, Princeton, NJ 08541-0001. Grades 8-Adult. Documentation is in: Bart, W. M. (1972). Construction and validation of formal reasoning instruments. **Psychological Reports**, 30, 663-670. These three instruments test formal thinking in the context of a subject matter. Examinees must apply rules of class or conditional logic to statements which are fictitious or contrary to fact in these deduction items.

Group Assessment of Logical Thinking (1982) by Vantipa Roadrangka, Russell H. Yeany and Michael Padilla, University of Georgia, Athens, GA 30602. Available from the authors. Grades 6-12. The test measures six formal operations: conservation, proportional reasoning, controlling variables, combinatorial reasoning, probabilistic reasoning and correlational reasoning. There are 21 multiple-choice science items which have follow-up questions to assess the student's reasons for choosing an answer. It is suitable for students with a sixth grade reading level. Total test reliability is .85.

Science Reasoning Level Test (n.d.) by Anna Dusynska. Available in: **Reasoning Level Test. Application of Piaget's Theoretical Model to the Construction of a Science Test for Elementary School** (ERIC No. ED 144988). Grades 3-6. This research instrument may be group administered within a single class period to rate stages of thinking using Piaget's categories of: preoperational, formal operational and concrete operational thinking. There are 16 multiple-choice questions which describe scientific experiments. The test was normed on Polish and American children.

Springs Task (1978) by Marcia C. Linn and Marian B. Rice, University of Chicago Laboratory School. Available from **Appendix to a Measure of Scientific Reasoning: The Springs Task**. (ERIC No. ED 163092). Also available in: Linn, M. C., & Rice, M. B. (1979, Spring). A Measure of Scientific Reasoning: The Springs Task, **Journal of Educational Measurement**, 16. Grades 5-Adult. An experiment is set up using springs and weights. Students are observed and interviewed as they learn the effects different weights have on the expansion of the springs.

Test of Logical Thinking (1979) by Kenneth G. Tobin and William Capie, University of Georgia, Athens, GA 30602. Available from the authors. Using situations from real-life, the authors test five formal operational concepts: controlling variables, proportions, combinations, probability and correlations. The questions are followed by choices of reasons for each response. There are two forms of this test. Internal consistency reliability for the total score is .85. Scores increased from grade 6-college. A factor analysis showed that all items related to one factor. The relationship between scores and clinical interviews was .82. The test appears to be a reasonable measure of formal reasoning.

Valett Inventory of Critical Thinking Abilities (1981) by Robert E. Valett. Available from Academic Therapy Publications, 20 Commercial Blvd., Novato, CA 94947-6191. The author's purpose is to evaluate the problem solving skills and abilities of children with learning problems. It is intended for use with ages 4-12, or older children experiencing learning problems. It is an individually administered performance test based on a neo-Piagetian model that emphasizes developmental stages. Tasks were chosen for the final form based on item response and content analysis. The authors report no evidence concerning the relationship of tasks to the constructs claimed to be measured or to student outcomes.

CREATIVITY TESTS

Make A Tree (1976) (no author) available from CTB/McGraw Hill, 2500 Garden Rd., Monterey, CA 93940. Grades PreK-1. This subtest in the CIRCUS series tests young children's divergent thinking ability. Children are asked to create two different trees, placing gummed stickers on a page. Pictures are scored for appropriateness, unusualness and difference between the two trees. There is no reliability information. The test was designed to minimize the need for verbal competence. There are good norms based on 2500 students and other good help with interpretation of results.

Pennsylvania Assessment of Creative Tendency (1968) by T. Jerome Rookey, Educational Improvement Center of Central N.J. Available from Tests in Microfiche (#008309) ETS Test Collection, Princeton, NJ 08541-0001. There are two, 39-item forms available of this survey of attitudes toward creativity, ambiguity and divergent thinking. It is appropriate for grades 4-9. Reliability estimates range from .79 to .92. The instrument seems to correspond well to other measures of creativity.

Possible Jobs (1963) by Arthur Gershon and J.P. Guilford. Available from Sheridan Psychological Services Inc., P. O. Box 6101, Orange, CA 92667. Grades 6-12. This brief test is one of several divergent thinking tests published by Sheridan where students are asked to generate ideas from given information. For this test, the prompts are emblems which represent a person's job. Students list up to six possible jobs for each emblem. They are scored for the number of appropriate responses. Internal consistency reliability is .70. Validity is based on factor analysis using Guilford's Structure of the Intellect categories.

Seeing Problems (1969) by Philip R. Merrifield and J. P. Guilford. Available from Sheridan Psychological Services Inc., P. O. Box 6101, Orange, CA 92667. Grades 7-12 and Adult. This is another brief divergent thinking test. It assesses a student's ability to conceptualize an object in terms of its properties and to infer potential problems with that object (for example, a candle drips wax, needs to be lit, may go out, etc.). Responses are analyzed according to Guilford's Structure of the Intellect category "cognition of semantic implications" which is the ability to plan well and foresee potential problems. The manual is brief and the scoring guide was not part of the specimen set. Reliability for the six-item form is about .67. Although the instrument has been used somewhat in research, its usefulness in educational setting has not been demonstrated.

Test of Creative Potential (1973) by Ralph Hoepfner and Judith Hemenway. Available from Monitor, P.O. Box 2337, Hollywood, CA 90078. Grades 2-12. This test is intended to measure the creativity factors of fluency, flexibility, originality and elaboration in Guilford's Structure of the Intellect model. However there are no subscores available for these factors. There are moderate correlations with intelligence and measured pre- and post-test differences in a program stressing creativity. The questions are open-ended prompts, using both language and non-language abilities. There is an interrater reliability of .76-.99. There are norms by grade level.

Test of Divergent Thinking, Test of Divergent Feeling and Williams Scale (1980) by Frank Williams. Available from D.O.K. Publishers, East Aurora, NY 14052. Grades 1-12. These three tests are part of the Creativity Assessment Packet, a series of instruments to assess a combination of cognitive and affective factors related to children's creative behavior. The **Test of Divergent Thinking** consists of 12 line prompts which the examinee makes into a picture. It is scored for fluency, flexibility, originality and elaboration which are based on the structure of intellect factors. The **Test of Divergent Feeling** asks students to self-report on their behavior. These self-reports are translated into inferences about how curious, imaginative, complex and risky the examinee is. The **Williams Scale** is an observational

checklist filled out by parents and teachers which covers the same eight creativity aspects as on the other two instruments. Test-retest reliabilities "were in the sixties." Correlation between ratings and other measures were .59 and .67 while correlation between parent and teacher ratings is .74. The former correlations seem somewhat low. There is little information on validity. There is extensive help with scoring.

Thinking Creatively With Sounds and Words (1973) by E. Paul Torrance, Joe Khatena and Bert F. Cunnington. Available from Scholastic Testing Service, 480 Meyer Rd., Bensenville, IL 60106. Grades 2-12 or Adult. There are two separate tests: **Sounds and Images** and **Onomatopoeia and Images** in this battery and two forms of both tests. A sound recording contains the narration for the test along with the sound and word prompts. Students are asked to write about each sound or word. A scoring key for rating responses is available with a possible four points for each item. Each test takes 30-35 minutes to administer.

Torrance Test of Creative Thinking (1974) E. Paul Torrance. Available from Scholastic Testing Service, 480 Meyer Rd., Bensenville, IL 60106. There are two tests: a verbal test called **Thinking Creatively With Words** (forms A and B) for grades 4-Adult; and a figural test called **Thinking Creatively with Pictures** (forms A and B) for grades K-Adult. Both tests are group administered though it is advised to contain the group to a normal class size. The verbal test could be administered individually to K-3. In the verbal test, examinees are asked to list possible questions, problems, improvements and uses of objects or persons in pictures. Responses are scored for fluency, flexibility, originality and elaboration, depending on the task. There are seven tasks. The Figural test has three visual tasks--involving constructing or completing a picture or a series of pictures. Drawings are scored for fluency, flexibility, originality and elaboration. Reliabilities are good and there has been extensive study of validity. Means are provided for various study groups.

ACHIEVEMENT TESTS

The following tests claim to include items which measure higher order thinking skills. A few provide a separate higher order thinking skills score either from rescoring items across subtests or by having a separate subtest. Some publishers will provide a score for an individual subtest but do not provide a single higher order thinking skills score across subtests. Some list the item numbers which test inferential, analytical or evaluative skills so that the user could compute scores on these variables if desired.

Assessment of Reading Growth (1980) available from Jamestown Publishers, P.O. Box 6743, Providence, RI 02940. Grades 3, 7, and 11. Literal and inferential comprehension are assessed in this reading survey taken from the National Assessment of Educational Progress released items. There are norms for the three levels for both literal and inferential comprehension.

California Achievement Test, Forms E and F (1985) available from CTB/McGraw Hill, 2500 Garden Rd., Monterey, CA 93940. Grades K-12. Items are cross-referenced to Bloom's Taxonomy. A higher order thinking skills score for grades 3-12 is available from the publisher and is derived from questions in the reading comprehension, language expression and mathematics concepts and applications subtests.

Comprehensive Tests of Basic Skills, Forms U and V (1981) available from CTB/McGraw Hill, 2500 Garden Rd., Monterey, CA 93940. Grades K-12. Category objectives are cross-referenced to Bloom's Taxonomy and individual test items are listed for each category. There are inference and evaluation questions in language arts, reading and mathematics (K-12), Science and Social Studies (2-12) and Reference Skills (3-12). The user must compute a higher order thinking skills score by analyzing items.

Iowa Test of Basic Skills: Early Primary and Primary Batteries and Tests of Achievement and Proficiency (high school) (1985) available from Riverside Publishing Co., 8420 Bryn Mawr Avenue, Chicago, IL 60631. Grades K-12. Items are coded by individual skill objectives. Individual responses are listed for: inferential meaning and predicting outcomes in the Listening subtest; inferring underlying relationships and developing generalizations in the Reading subtest; inferring behavior and living conditions and interpreting and relating data from the Maps subtest; and classification in the Reference Materials subtest. The user could compute HOTS scores based on the item statistics; the publisher doesn't provide such scoring.

Metropolitan Achievement Tests, 6th edition, Form L (1985) available from Psychological Corporation/Harcourt Brace Jovanovich, 555 Academic Court, San Antonio, TX 78204-0952. Grades K-12. A raw score for higher order thinking skills is available from the publisher. This score is translated to "low", "average", or "high" for comparison. The score is derived from reading, math, science and social studies questions which test the higher levels of Bloom's Taxonomy.

The National Tests of Basic Skills (1985) available from American Testronics, P.O. Box 2270, Iowa City, Iowa 52244. Grades Pre-school-College. Percent correct scores are available for individual skills objectives. The Reading Comprehension subtest includes inferential and evaluative comprehension which may be scored separately by the publisher.

Reading Yardsticks (1981) available from Riverside Publishing Co., 8420 Bryn Mawr Ave., Chicago, IL 60631. Grades K-8. Separate scores for Interpretive reading (K-8) and Evaluative reading (3-8) are available.

Scan-Tron Reading Tests (1985) available from SCAN-TRON Corporation, Reading Test Division, 2021 East Del Amo Blvd., Rancho Dominiquez, CA 90220. Grades 3-8. Information is included about the skills measured by each item so a user could compute a subscore for Inferential Comprehension. Average percent correct is available for comparison purposes. SCAN-TRON does not provide scoring at this level.

SRA Achievement Series (1978-1985) available from Science Research Associates, Inc., 155 Wacker Drive, Chicago, IL 60606. Grades K-12. Item objective information is included so the user could compute scores for perceiving relationships, drawing conclusions and understanding the author in the Reading Comprehension subtest (K-10); identifying insufficient or extraneous information in Math and Word Problems (5-12); interpreting visual materials and determining consequences in Social Studies (5-12); and applying scientific inquiry methods in Science (5-12). Norms are derived from equating items with the **Survey of Basic Skills**.

Stanford Achievement Test, Forms E and F (1982) available from Psychological Corporation/Harcourt Brace Jovanovich, 555 Academic Court, San Antonio, TX 78204-0952. Grades 1-9. Items are cross-referenced to specific objectives and individual scores by content cluster are available from the publisher's scoring service. The "Using Information" cluster (grades 3-9) rescores items from several topical areas and may be an indicator of higher order thinking, though it is not promoted as such by the publisher.

Stanford Achievement Test, Five Edition (1982) available from Psychological Corporation/Harcourt Brace Jovanovich, 555 Academic Court, San Antonio, TX 78204-0952. Grades 1-9. This edition rescores the SAT (1982) to include a higher order thinking skills score which uses questions from each of the content areas. There are 1985-1986 norms available from the publisher.

Stanford Test of Academic Skills, Forms E and F (1982) available from Psychological Corporation/Harcourt Brace Jovanovich, 555 Academic Court, San Antonio, TX 78204-0952. Grades 8-College. The "Using Information" objective cluster score includes inquiry skills from Science and Social Science, and reference skills from the English subtest. This score could be used as a higher order thinking skills score.

Survey of Basic Skills, Forms P and Q (1985) available from Science Research Associates Inc., 155 Wacker Drive, Chicago, IL 60606. Grades K-12. The Individual Skills Profile report lists skill objectives for each student. Listening Comprehension includes inference questions (K-1); Reading Comprehension includes inference and analysis (1-12); Mathematics includes problem solving skills (1-12); Social Studies includes interpretation and reasoning skills (4-12); and Science includes inquiry skills (4-12).

The Three-R's Tests, Forms A and B (1982) available from Riverside Publishing Co., 8420 Bryn Mawr Ave., Chicago, IL 60631. Grades K-12. Item analysis scores for logical relationships (grades 2-12), literary analysis and author's purpose (grades 3-12) are available from the publisher.

ABILITY TESTS

Developing Cognitive Abilities Test (1980) by John W. Wick and Jeffrey K. Smith. Available from American Testronics, P.O. Box 2270, Iowa City, IA 52244. Grades 2-12. This test claims to assess all the levels of Bloom's Taxonomy but it is heavily weighed toward knowledge, comprehension and application in levels 3-8. Nearly half of the items for levels 9-12 include analysis and synthesis questions.

Structure of Intellect Learning Abilities Test (SOI-LA) (1985) by Mary Meeker, Robert Meeker and Gale H. Roid. Available from Western Psychological Services, 12031 Wilshire Blvd., Los Angeles, CA 90025. Grades K-Adult. Based on Guilford's multifactor approach to intelligence, the authors devised a two-three hour test to diagnose student abilities. It may be individually or group administered. There are two alternative forms measuring 26 basic abilities and additional forms for arithmetic, reading, gifted screening, primary and reading readiness based on subsets of the 26 abilities. The 26 subtests sample from the 120 in the Structure of the Intellect model. The instrument has undergone extensive development. There are norms and assistance with interpretation and use.

To be published

A multi-dimensional school ability test for K-12 based on Robert J. Sternberg's Triarchic Theory of Intelligence is scheduled to be published in 1989 by the Psychological Corporation.

OBSERVATION SCALES

These instruments can be used to assess the process of instruction with regard to promoting thinking.

Florida Taxonomy of Cognitive Behavior (1968) by Bob B. Brown, Richard L. Ober, Robert S. Soar and Jeaninne N. Webb, University of Florida. Available from Tests in Microfiche (#005949) ETS Test Collection, Princeton, NJ 08541-0001. This classroom observation instrument lists 55 teacher and student behaviors which are organized by Bloom's Taxonomy. Directions for scoring each behavior are given but no assistance in interpreting data is provided.

Stallings-Simon Observation Instrument (n.d.) by Jane Stallings and Sandra Simons. Available from: Sandra Simons, Educational Consultant, 2606 Spring Blvd., Eugene, OR 97403. This classroom observation instrument is intended to provide information to teachers so they can better promote student thinking. It focuses on only one aspect of classroom processes: instructional interactions between the teacher and students, especially questioning. There is some information on how to score results. Full use of the instrument requires training.

A Thinking Skills Teaching Inventory (1985, January) Barry K. Beyer, George Mason University. Available in: Teaching Thinking Skills: How the Principal Can Know They are Being Taught in **NASSP Bulletin**. School Administrators may use this checklist to survey the status of higher order thinking in the classroom. The instrument is essentially an outline and has not been cast in the format of a formal observation tool.

APPENDIX B

SUMMARY TABLE OF

HIGHER ORDER THINKING SKILLS TESTS

HOTS INSTRUMENTS

Summary Table of Instrument Characteristics

INSTRUMENT	FOCUS	GRADES	SUBJECT SPECIFICITY	NO. FORMS	NO. LEVELS	NO. ITEMS	ITEM TYPE	ADM. TIME
CRITICAL THINKING AND PROBLEM SOLVING:								
Applications of Generalizations Test (1969)	Generalizations	4-12	Social Studies	1	1	65	M.C.	?
Cornell Class Reasoning Test (1964)	Class Reasoning	4-12	General	1	1	72	M.C.	40 min.
Cornell Conditional Reasoning Test (1964)	Conditional Reasoning	4-12	General	1	1	72	M.C.	40 min.
Cornell Critical Thinking Tests	Critical Thinking	X 4-12 Z adult	General	1		71 52	M.C.	60 min.
Ennis-Weir Critical Thinking Essay Test (1985)	Critical Thinking	9-adult	General	1	1	1	Essay	70 min.
Judgment: Deductive Logic & Assumption Recognition (1971)	Critical Thinking	7-12	General	1	1	135	M.C.	135 min.
Means-Ends Problem Solving (1975)	Interpersonal Problem Solving	5-7 9-adult	General	1	2	10	Essay, Interview	None given
New Jersey Test of Reasoning Skills (1985)	General	4-adult	General	1	1	50	M.C.	30-60 min.
Primary Test of Higher Processes Thinking (1978)	Bloom's Taxonomy	2-4	General	1	1	55	M.C., open, match	60 min.

Scoring	Norms	Other Interp.	Reliability	Validity	Comments	Availability
Hand	Fair	None	Fair	Some+	Specifically tied to Taba curriculum	Tests in Microfiche (#0084261), ETS Test Collection, Princeton, NJ 08541-0001
Hand	Fair	None	Fair	Some	Structured formal logic only.	Illinois Critical Thinking Project, Univ. of Illinois Champaign, IL 61820 or ERIC ED 003818
Hand	Fair	None	Fair	Some	Structured formal logic only.	Illinois Critical Thinking Project, Univ. of Illinois Champaign, IL 61820 or ERIC ED 003818
Hand, Machine	Fair	Some	Fair-Good	Some		Midwest Publications P.O. Box 448 Pacific Grove, CA 93950
Hand	Fair	Some	Fair-Good	Some	Scoring requires training.	Midwest Publications P.O. Box 448 Pacific Grove, CA 93950
Hand	None	None	No Info.	None	Has five separate aspect specific tests.	IOX Assessment Associates Box 24095 Los Angeles, CA 90024
Hand	Fair	Some	Good-Excellent	Extensive	Scoring and administration requires training.	Department of Mental Health Science, Hahnemann Univ., 112 N. Broad St. Philadelphia, PA
Machine only	Fair	None	Good-Excellent	Some	No reliabilities are reported for sub-skills. Test is rented from publisher. Tied to the Philosophy for Children Program.	Institute for the Advancement of Philosophy for Children, Montclair State College, Upper Montclair, NJ 07043
Hand	Fair	None	Fair	None		Tests in Microfiche (#013161), ETS Test Collection, Princeton, NJ 08541-0001

INSTRUMENT	FOCUS	GRADES	SUBJECT SPECIFICITY	NO. FORMS	NO. LEVELS	NO. ITEMS	ITEM TYPE	ADM. TIME
CRITICAL THINKING AND PROBLEM SOLVING (continued...)								
Purdue Elementary Problem Solving Inventory	Problem Solving	2-6	General	1	1	49	M.C.	40-50 min.
Ross Test of Higher Cognitive Processes (1976)	Bloom's Taxonomy	4-6	General	1	1	105	M.C.	140 min.
TAB Test: An Inventory of Science Methods (1966)	Problem Solving	4-6	Science	2	1	2	Performance situations	?
Test of Enquiry Skills (1979)	Critical Thinking	7-10	Science	1	1	87	M.C.	1½ - 5 hours depending on grade
Test of Science Comprehension (1963)	Interpreting Data	4-6	Science	1	1	30	M.C.	90 min.
Test on Appraising Observations (1983)	Appraising Observations	7-adult	General	1	1	50	M.C.	50 min.
Think It Through (1976)	Problem Solving	preK-1	General	1	2	31	M.C.	30-40 min.
Watson-Glaser Critical Thinking Appraisal (1980)	Critical Thinking	9-adult	General	2	1	80	M.C.	40 min.

Scoring	Norms	Other Interp.	Reliability	Validity	Comments	Availability
Hand	Fair	None	Fair	Some	Uses a filmstrip and tape to present questions.	Gifted Education Resource Institute, Purdue Univ., S. Campus Courts, Bldg. G, W. Lafayette, IN 47907
Hand, Machine	Good	Some	Excellent	Some	Scoring can be time consuming	Academic Therapy Publications, 20 Commercial Blvd. Novato, CA 94947-6191
?	Fair	Some	Poor	Some	Student chooses sequence of "experiments" to answer a question by pulling tabs to uncover results.	Tests in Microfiche (#007741), ETS Test Collection, Princeton, NJ 08541-001
Hand	Fair	Some	Fair	Some	Only one subtest pertains to critical thinking.	Australian Council for Educational Research, Frederick St., Hawthorne, Victoria 3122, Australia
Hand	Fair	Some	Fair	Some		A Test of Science Comprehension for Upper Elementary Grades, Science Ed. 47, 319-320 (1963).
Hand	Fair	Some	Poor	Extensive		Institute for Educational Research and Development, Memorial University of Newfoundland, St. Johns, Newfoundland, Canada A1B3X8
Hand,	Good	Extensive	Fair	Some	Questions are read to students.	CTB/McGraw Hill Del Monte Research Park, 2500 Garden Road, Monterey, CA 93940
Hand, Machine	Good	Some	Fair-Good	Extensive		Psychological Corporation, 555 Academic Court, San Antonio, TX 78204-0952

Instrument	Focus	Grades	Subject Specificity	No. Forms	No. Levels	No. Items	Item Type	Adm. Time
DEVELOPMENTAL TESTS:								
Arlin Test of Formal Reasoning (1984)	Piaget	6-adult	General	1	1	32	M.C.	45 min.
Classroom Test of Formal Reasoning (1978)	Piaget	8-12	General	1	1	15	M.C.	75-100 min.
Formal Operations Measure (no date)	Piagnet	Adult	Science	2	1	7	Open-ended	45-60 min.
Formal Operations Test	Piaget	8-adult	Biology History Literature	1 each	1 each	30 each	M.C.	?
Springs Task (1978)	Piaget	5-adult	General	1	1	19	Indiv. Open-ended	15 min.
Test of Logical Thinking (1979)	Piagnet	6-adult	General	2	1	10	M.C.	38 min.
Valett Inventory of Critical Thinking Abilities (1981)	neo-Piagetian	4-12	General	1	1	*	Open-ended	N.A.
Understanding in Science (1975)	Piaget	7-9	Science	1	1	24	M.C., Short resp.	40 min.

Scoring	Norms	Other Interp.	Reliability	Validity	Comments	Availability
Hand, Machine	Fair	Some	Poor-	Some	Follow-up questions ask for reasoning behind answer. Hand scoring is awkward.	Slosson Education Publications, P.O. Box 280 E. Aurora, NY 14052
Hand	Fair	Some+	Fair	Some+		The Development and Validation of a Classroom Test of Formal Reasoning, Journal of Res in Science Teaching, 5, 11-24 (1978)
Hand	None	Some	None	None		Tests in Microfiche (#010271), ETS Test Collection Princeton, NJ 08541-0001
Hand	Fair	None	None	Some	Three separate tests.	ETS Test Collection, Tests in Microfiche (8422, 8423, 8424) Princeton, NJ 08541-0001
Hand	?	?	Good	Some	Requires an apparatus.	A Measure of Scientific Reasoning: The Springs Task. J. Ed. Measurement, 16, 1978. ERIC ED 163092
Hand	Fair	None	Good	Some+	Examinees pick answer and justification	Kenneth Tobin and William Copie, U. of Georgia Athens, GA 30602
Hand	None	Some	None	None	Children complete tasks until child misses 4 out of 5 in a row.	Academic Therapy Publications, 20 Commercial Blvd. Novato, CA 94947-6191
Hand	None	Some	None	Some		Australian Council for Educational Research, Ltd., Frederick St., Hawthorne Victoria 3122, Australia

Instrument	Focus	Grades	Subject Specificity	No. Forms	No. Levels	No. Items	Item Type	Adm. Time
CREATIVITY TESTS:								
Make a Tree (1976)	Divergent	preK-1	General	1	1	2	Open-ended	Not provided
Pennsylvania Assessment of Creative Tendency (1968)	Affective Correlates of Creativity	4-9	General	2 short 2 long	1	19 short 39 long	Likert	Not provided
Possible Jobs (1963)	Divergent Thinking	6-12	General	2	1	3	Open-ended	10 min.
Seeing Problems (1969)	Sensitivity to problems	7-adult	General	2	1	6	Open-ended	7 min.
Test of Creative Potential	Divergent thinking	2-12	General	2	1	3	Open-ended	30 min.
Test of Divergent Thinking (1980)	Divergent thinking	3-12	General	1	1	12	Open-ended	20-25 min.
Test of Divergent Feeling (1980)	Affect	3-12	General	1	1	50	M.C.	20-30 min.
Williams Scale (1980)	Creativity General	1-12	General	1	1	48	Check-list	30 min.
Thinking Creativity with Sounds and Words (1973)	Creativity-General	3-adult	General	2	2	8	Open-ended	30-35 min.
Torrance Test of Creative Thinking (1970)	Divergent thinking	K-adult	Science	2	1	10	Open-ended	1 hr. 45 min.

Scoring	Norms	Other Interp.	Reliability	Validity	Comments	Availability
Hand	Good	Some+	None	None	Part of the CIRCUS Test Battery, Lots of help with interpretation	CTB/McGraw Hill 2500 Garden Road Monterey, CA 93940
Hand	Fair	Some	Good	Some+	Long and short forms of the survey are available. Short forms should only be used experimentally.	Tests in Microfiche (#8309), ETS Test Collection Princeton, NJ 08541-0001
Hand	Fair	Some	Fair	Some+		Sheridan Psychological Services P.O. Box 6101 Orange, CA 92667
Hand	Fair	Some	Fair	Some+	Although this seems to represent the factor claimed, usefulness in all educational settings has not been demonstrated.	Sheridan Psychological Services P.O. Box 6101 Orange, CA 92667
Hand	Fair	None	Fair-Good	Some		Monitor, P.O. Box 2337 Hollywood, CA 90078
Hand	Fair	Some	Poor	Some	Part of the Creativity Assessment Packet	DOK Publishers E. Aurora, NY 14052
Hand	Fair	Some	None	Some	Part of the Creativity Assessment Packet	DOK Publishers E. Aurora, NY 14052
Hand	Fair	Some	None	Some	Part of the Creativity Assessment Packet	DOK Publishers E. Aurora, NY 14052
Hand	?*	?*	Interrater-Excellent	?*	Uses a recording for sounds. Some information in a technical manual which we did not get.*	Scholastic Testing Service, 480 Meyer Road Bensonville, IL 60106
Hand	Good	Some	Good	Extensive	Based on a broad definition of the creative act.	Scholastic Testing Service, 480 Meyer Road Bensonville, IL 60106

*Technical manual not available.

TEST	FOCUS	GRADES	SUBJECT AREAS	SCORE OBTAINED	AVAILABILITY
ACHIEVEMENT TESTS:					
Assessment of Reading Growth (1980)	Inferential Comprehension	3,7,11	Reading	Inferential comprehension score available	Jamestown Publishers P.O. Box 6743 Providence, RI 02940
California Achievement Test (1985)	Bloom's Taxonomy	K-12	Reading Language	HOTS score re-scored from other subtests by publisher	CTB/McGraw Hill 2500 Garden Road Monterey, CA 93940
Comprehensive Tests of Basic Skills (1981)	Bloom's Taxonomy	K-12	Reading, Math Language, Science, Social Studies, Ref. Skills	Items are cross-referenced to Bloom's Taxonomy. User must generate HOTS score.	CTB/McGraw Hill 2500 Garden Road Monterey, CA 93940
Iowa Test of Basic Skills (1985)	Various depending on subtest	K-12	Listening, Reading, Maps, Ref. materials	Same as above.	Riverside Pub. Co. 8020 Bryn Mawr Avenue Chicago, IL 60631
Metropolitan Achievement Tests (1985)	Bloom's Taxonomy	K-12	Reading, Math Science, Social Studies	HOTS score re-scored from other subtests by publisher	Psychological Corp. 555 Academic Court San Antonio, TX 78204
National Tests of Basic Skills (1985)	Inference Evaluation	PreK-adult	Reading	Inferential and evaluative comprehension aggregated by publisher; % correct on individual objectives.	American Testronics P.O. Box 2270 Iowa City, IA 52244
Reading Yardsticks	Interpretation Evaluation	K-8	Reading	Interpretive and evaluative reading available from publisher	Riverside Pub. Co. 8020 Bryn Mawr Avenue Chicago, IL 60631
Scan-Tron Reading Tests (1985)	Inferential Comprehension	3-8	Reading	Items are cross-referenced to skills. User must generate score.	SCAN-TRON Corporation Reading Test Division 2021 East Del Amo Bv. Rancho Dominiquez, CA 90220
SRA Achievement Series (1985)	Various, depending on subtest	K-12	Reading Math Social Studies Science	Items are cross-referenced to skills. User must generate score.	Science Research Associates, Inc. 155 Wacker Drive Chicago, IL 60606
Stanford Achievement Test (1981)	Using information	1-9	English Science Social Studies	Using information score, re-scored from other subtests by publisher.	Psychological Corp. 555 Academic Court San Antonio, TX 78204-0932
Stanford Test of Academic Skills (1982)	Using information	8-adult	English Science Social Studies	Using information score, re-scored from other subtests by publisher.	Psychological Corp. 555 Academic Court San Antonio, TX 78204-0932
Survey of Basic Skills	Various, depending on subtest	K-12	Listening, Reading, Math Social Studies Science	Users consult the skills profile report for each subtest	Science Research Associates, Inc. 155 Wacker Drive Chicago, IL 60606
The Three R's Tests (1982)	Logical relationship Literary analysis Author's purpose problem solving	K-12	Reading Language Arts Mathematics	Item analysis scores available from publisher	Riverside Pub. Co. 8420 Bryn Mawr Ave. Chicago, IL 60631

Interpretation of Table Codes

Norms (Value judgement implied)

None	No normative information is provided
Fair	Has some standards of comparison, e.g., means of research sample, decile norms or item statistics.
Good	Has norms based on a good sized sample or lots of other information.
Excellent	Has norms based on a national sample, and other information.

Other Interpretation (No value judgement as to the quality of the assistance is implied)

None	No help with interpretation provided.
Some	Has some help with interpreting scores, e.g., what the various scores mean.
Some+	Has information on what the scores mean and some help with use in instruction.
Extensive	Has extensive information on what the scores mean and how to use them in instruction.

Reliability (Value judgement implied)

None provided	No information was found.
Poor	All r's below .70
Fair	At least one reported r is greater than .70
Good	Total r is greater than .85; most subtests have r greater than .75.
Excellent	Several kinds reported; total score r is greater than .90; most subtest scores greater than .80

Validity (This describes the quantity of information available, not necessarily the extent to which the instrument is valid.)

No information	No information on validity is reported.
Some information	At least one activity related to validation is reported.
Some+ information	Validity was examined in several different ways.
Extensive information	Special effort was made to examine validity and there is a large research base on the instrument.

APPENDIX C

TESTING RESOURCES

LOCAL, STATE AND FEDERAL DEPARTMENT OF EDUCATION PROGRAMS

California

The California State Department of Education is in the process of explicitly incorporating thinking skills in the curriculum and in its statewide assessment. The **Survey of Academic Skills**, produced by the California Assessment Program, includes at least 40 percent critical thinking questions on the tests of History-Social Science, Mathematics, Language Arts and Science. Math and Language Arts are currently administered to grades 3, 6, 8, and 12 (300,000 students per grade). There is also a separate Writing Sample that requires students to evaluate, solve problems and speculate. The History-Social Science section was administered first in 1985. The Science section was administered in 1986 for grade 8. The Writing Assessment will be first administered in 1987 to all 8th graders.

Contact: Pete Kneedler, California Assessment Program, California State Department of Education, 721 Capitol Mall, Sacramento, CA 95814-4785.

Connecticut

The Connecticut State Department of Education, with the assistance of several experts in the field (such as Robert Ennis, Robert Sternberg and Edys Quellmalz), is incorporating critical thinking skills into it's grade 4 tests in mathematics, language arts, reading and listening. There is also a writing sample. Each subject includes objectives which reflect critical thinking skills derived from Ennis and Sternberg.

Contact: Joan Baron, Conn. State Department of Education/Office of Research and Evaluation, P. O. Box 2219, Hartford, CT 06145

Illinois

The State Department of Education contracted with the Center for the Study of Reading at the University of Illinois, Champaign-Urbana to develop a statewide test to assess the reading ability of 3rd, 6th, 8th and 10th grade students in ways more consistent with current research on reading, as expressed in **A Nation of Readers**. This view states that prior knowledge is an important determinant of comprehension, one needs a complete story to have structural and topical integrity, and good readers ask questions of text as they go. The pilot form of the test included four types of questions designed to assess various aspects of this conception of the reading process. They are: a) prior knowledge or topic familiarity; b) comprehension; c) meta-cognitive skills such as sensitivity and flexibility; and d) habits and attitudes. Since this approach is intended to measure the sustained effort of students to comprehend what they read and use appropriate reading strategies for the type of material read, the developers consider it a measure of HOTS.

Comprehension questions make use of new formats. Besides the standard multiple-choice questions, there are multiple-multiple choice (more than one correct answer), score every answer (rating), and question selection (choose good questions to ask). Preliminary data have shown that as students become more sophisticated, their skills in each of the subtests become better. The first administration of this test is scheduled for 1988.

Contact: Carmen Chapmen, Illinois State Department of Education. 100 N. First St., Springfield, IL 62777.

Michigan

The Michigan Educational Assessment Program is developing a long range plan which includes expansion to new subject areas and grade levels, as well as testing a broader conceptual range of skills. Building on the annual assessment of all fourth, seventh and tenth grade students in Reading and Mathematics, new tests in these areas will include thinking skills and a broader conceptual range (beyond knowledge). In addition, tests in Health, Science, Career Development Social Studies and Writing have been developed, some for grades 4, 7 and 10, others for grades 5, 8 and 11; these tests will be offered on a voluntary basis. A common definition of thinking is being developed and will be incorporated in future assessments in each of these areas.

Contact: Edward D. Roeber, Michigan Educational Assessment Program. Michigan Department of Education, P.O. Box 30008, Lansing, MI 48909.

National Assessment of Educational Progress

NAEP cyclically assesses 9, 13 and 17 year olds and adults in the areas of art, music, reading, science, and social studies. These assessments include critical thinking or problem solving questions and situations as part of the subject matter tests. Released items have shown up on other achievement tests such as the **Assessment of Reading Growth**.

Contact: Educational Testing Service, Princeton, NJ 08541-0001.

Pennsylvania

The Pennsylvania Department of Education's Educational Quality Assessment program began in 1970. In 1979, EQA revised its objectives to include "analytical thinking" which was defined as information management, logical thinking, problem solving, and decision-making. The items developed to test analytical thinking were divided into several forms and, as with the rest of EQA, they were administered using matrix-sampling.

Analytical thinking questions are built around problem situations which interest the students in the grades being assessed. A single passage prompts questions which test inference, information processing and decision-making or drawing conclusions. The test constructors followed up the pilot test with interviews to be certain that students were choosing answers for the right reasons. The EQA with the analytical thinking subtest has been administered to students in grades 5, 8 and 11 in 1985 and grades 4, 6, 7, 9 and 11 in 1986 using matrix sampling.

Contact: James R. Masters, Educational Quality Assessment, Pennsylvania Department of Education, Box 911, Harrisburg, PA 17126.

The Pittsburgh Public Schools Monitoring Achievement in Pittsburgh (MAP) developed the **Critical Thinking Test** (1983) to assess students critical thinking abilities in the social

sciences using an essay test. Students read a prose passage related to the social studies curriculum and then write an essay which evaluates or draws inferences from what they read. Essays are scored for: topic statement, evidence, explanations, concluding statement, organization and response to task. Raters had agreement within one point for 96 percent to 98 percent of the essays.

Contact: Division of Curriculum Development, 341 South Belleville Ave., Pittsburgh, PA 15213.

COLLEGES, UNIVERSITIES, AND PROFESSIONAL ORGANIZATIONS

American Federation Of Teachers, 555 New Jersey, Ave., NW, Washington, DC 20001 (202)879-4400. The AFT has a Critical Thinking Project. Marilyn Rauth is the Executive Director and Debbie Walsh is the Assistant Director. Some of their activities include publishing a book called **Critical Thinking: From Educational Ideal to Educational Reality**--an overview of the critical thinking movement; Training-of-Trainers Program; Critical Thinking Network; a videotape called **Inside Your Schools** which looks at teaching for thinking; and a survey of states on state level activities related to critical thinking.

Association for Supervision and Curriculum Development, 225 N. Washington Street, Alexandria, VI, (703)549-9110. The ASCD has done a lot in encouraging the teaching of thinking. Some examples: Numerous articles and several special editions in the organization's journal **Educational Leadership**; organization of a group of national organizations to encourage the development of a thinking perspective among the affiliated groups. For more information on the Collaborative on Teaching Thinking, contact Dr. Ronald Brandt, Executive Editor, ASCD, 125 N. West St., Alexandria, VA 22314; and publication of a book called **Developing Minds--A resource book for teaching thinking**.

University of Massachusetts at Boston, Harbor Campus, Boston, MA 02125-3393 (617)929-7900. This campus offers a Masters of Arts Degree in Critical and Creative Thinking.

Center for Critical Thinking and Moral Critique, Sonoma State University, Rohnert Park, CA 94928, (707)664-2940. This center has a yearly conference on critical thinking and educational reform.

BIBLIOGRAPHIES OF HIGHER ORDER THINKING SKILLS ASSESSMENT TOOLS

Creativity and Divergent Thinking (1986) Princeton, NJ: Test Collection, Educational Testing Service. Approximately 75 tests of creative thinking or divergent thinking are described in this bibliography. Subtest scores are listed and availability information is included.

Critical Thinking Tests (1986) Robert H. Ennis, Champaign,IL: Illinois Critical Thinking Project, University of Illinois, 1310 S. Sixth Street, Champaign, IL 61820. Brief reviews of major reasoning tests currently available. Includes seven general and four aspect-specific critical thinking tests.

The Ninth Mental Measurements Yearbook (2 volumes) (1985) James V. Mitchell, Jr., (Ed.). Lincoln, NE: Buros Institute of Mental Measurements. The classification scheme in this grandparent of test reviewing sources does not include "critical thinking" but it is possible to look under a known title for an in-depth review of a test or to use the "score index" to look for tests producing scores with various labels. The following scores have been indexed: critical comprehension, critical thinking, creative thinking, divergent thinking, logical ability, logical/analytical, and problem solving.

Reasoning, Logical Thinking and Problem Solving (1986) Princeton, NJ: Test Collection, Educational Testing Service. Abstracts and availability information for 133 tests are included in this bibliography. The majority of tests are aptitude measures, though three are critical thinking tests included as well.

Role of Performance Assessment in Tests of Problem Solving (1981) Thomas P. Sachse. Portland, OR: Clearinghouse for Applied Performance Testing, Northwest Regional Educational Laboratory. Reviews of 13 school ability, life-skills, problem solving and critical thinking tests are included in this bibliography. Six features of each test are examined: definitions of problem solving, measurement strategy, performance assessment, reliability, uses of the test, and validity.

Testing for Critical Thinking: A Review of the Resources (1979) Bruce L. Stewart, Champaign, IL: Rational Thinking Reports No. 2, Illinois Rational Thinking Project, Univ. of Illinois-Urbana. (ERIC No. ED 183588). Twenty-five critical thinking tests are reviewed. Information on reliability and validity as well as item analysis is included. Most of the tests have a pre-1970 copyright, yet a few are still available in updated editions.

Tests: a Comprehensive Reference for Assessments in Psychology, Education and Business (1983) and **Tests Supplement** (1984) Richard C. Sweetland and Daniels J. Keyser (Eds.). Kansas, City: Test Corporation of America. This bibliography lists ordering and subtest information without evaluating instruments. Critical thinking tests may be found under the headings for English, Achievement and Aptitude, and Gifted. There is no cross-reference to subtest scores.

Tests in Print III (1983) James V. Mitchell (Ed.). Lincoln, NE: Buros Institute of Mental Measurements. A companion volume to **The Mental Measurements Yearbook** (MMY). Lists availability and price information for 2672 tests, including most of the tests in MMY. The same classification scheme as that in MMY is used, and the subtests are not indexed.

BOOKS AND ARTICLES

Costa, A.L. (1983, October 28) **"Thinking: How do we know students are getting better at it?"** Unpublished paper, California State University at Sacramento. This article describes a record keeping system for assessing growth in intellectual behavior which includes perseverance, planning, flexibility, awareness of own thinking, checking for accuracy, problem posing and applying knowledge and experience.

Costa, A.L. (Ed.). (1985) **Developing minds: A resource book for teaching thinking.** ASCD Publications, 225 N. Washington St., Alexandria, VA, (703)549-9110. This book is a collection of articles covering the topics of definitions, developing a curriculum for teaching thinking, assessing growth in thinking and the role of computers. The book also reviews 17 instructional programs for teaching thinking, has abstracts for a selection of readings in a broad variety of fields having to do with thinking, abstracts several critical thinking tests, and lists newsletters and professional organizations dealing with thinking.

Ennis, R.H. (1987) A taxonomy of critical thinking disp[illegible]ions and abilities. In J.B. Baron and R.J. Sternberg, **Teaching thinking skills: The[illegible] and practice** (pp. 9-26). New York: W. H. Freeman. This chapter is a good review of current issues with respect to assessing critical thinking.

Gough, H.G. (1985) "Free response measures and their relationship to scientific creativity." **Journal of Creative Behavior**, 19, 229-240. This paper describes 10 creativity measures used with adults in a study of scientific creativity.

Norris, S.P. and King, R. (1984) **"The design of a critical thinking test on appraising observations."** Institute for Educational Research and Development, Memorial University of Newfoundland. This report presents detail on how the authors went about developing and validating the Test of Appraising Observations. Included is their procedure for interviewing students in order to come up with an independent measure of quality of reasoning.

Stiggins, R.J., Rubel, E. and Quellmalz, E. **"Measuring thinking skills in the classroom: A teacher's guide."** Northwest Regional Educational Laboratory, 101 S.W. Main Street, Portland, OR 97204. This publication address how to assess HOTS in the classroom and how to embed HOTS skills into everyday lesson plans.

NEWSLETTERS

ASAP Notes. This newsletter is published by the Association of State Assessment Programs. It often contains articles on assessing HOTS.

Philosophy for Children Newsletter, The First Mountain Foundation, P.O. Box 196, Montclair, NJ 07042. This newsletter shares information about developments with the Philosophy for Children curriculum materials and the associated test the **New Jersey Test of Reasoning Skills**.

Cogitare, ASCD Network on Teaching Thinking, c/o John Barell, Montclair State College, Upper Montclair, New Jersey 07043.

Critical Thinking Network Newsletter, American Federation of Teachers, 555 New Jersey Ave., NW, Washington, CD 20001, (202)879-4400.

CURRICULUM REVIEWS

Costa, A.L. (Ed.). (1985) **Developing minds: A resource book for teaching thinking**. ASCD Publications, 225 N. Washington St., Alexandria, VA, (703)549-9110. One section of this book contains descriptions of 16 curriculum packages--SOI, Feuerstein, Strategic Reasoning, CoRT, Project IMPACT, Philosophy for Children, The California Writing Project, Future Problem Solving,Guided Design, Odyssey, Learning to Learn, Creative Problem Solving, Great Books, Building Thinking Skills, HOTS, and BASICS.

Marzano, R.J. (1986, September) "Practicing theory." **Cogitare**. Newsletter of the Thinking Skills Network sponsored by ASCD. This short article classifies several instructional programs into the categories of "structured formal logic," "informal logic," and "dialectic."

Sternberg, R.J. (1984, Sept.) "How can we teach intelligence?" **Educational Leadership**, 38-48. The author reviews three curriculum packages that he feels can be used to teach components of intelligence --Feuerstein, Philosophy for Children and Chicago Mastery Learning Reading Program.

Thinking skills academy. (1985) Workshop materials from Research for Better Schools, 444 N. 3rd St., Philadelphia, PA 19123. One section of the workshop materials review training programs--LAPS, Direct Instruction (Beyer), Feuerstein, Project Impact, and CoRT.

APPENDIX D

CHECKLIST FOR SELECTING A HIGHER ORDER THINKING SKILLS TEST

Checklist for Selecting a Higher Order Thinking Skills Test

I. Usefulness

A. Information Obtained

1. Do the stated uses of the instrument match up with what you want to use the information for?

2. Does the instrument or method measure the HOTS skills on which you want information?

3. Does the instrument assist with interpretation of results? Does it have criteria by which to judge results? This includes statements about what performance should be like at various grade levels. It could also include norms.

4. Is there information about how to use the results to plan instruction for students?

B. Logistics

1. Is the instrument or method easy to use?

2. Is it easy to score and interpret the results?

3. Is the length of time required to collect information acceptable?

C. Cost

1. Are costs within available resources? (Include costs of obtaining the instrument or method, training data collectors and collecting data.)

II. Technical Adequacy

A. Theoretical Basis

1. Do the supporting materials for the instrument or method present a clear definition of the aspects of HOTS that it claims to measure? Does the test manual discuss how this definition was developed and why the test has the content it has? Is evidence provided (based on research or theory) that the definition(s) and test content are reasonable?

B. Reliability

1. Was the instrument pilot tested?

2. Is there some measure of reliability available for the instrument?

a. For a structured-format test this includes at least item discriminations, internal consistency and test-retest reliabilities.

b. For an open-ended test this would include estimates of reliability of scoring such as interrater reliability.

c. If the results are going to be used to make important (and hard to reverse) decisions about individual students, reliability should be above .90. For group uses, or for educational decisions that are easily reversible, reliabilities should be above .75.

C. Validity: Is there evidence that the instrument measures what it claims to measure? Validity is in the relationship between the instrument and its use. There should be evidence that the instrument can be validly used for the purposes stated. For example, what evidence is there that the item types used measure the skill area?

1. For structured-format instruments an ideal set of validity studies would include:

a. The respondent understands what is being asked. Vocabulary or concepts unfamiliar to a group would make the instrument unusable for that group. This information would most likely be obtained by observing or interviewing students.

b. Right answers are only arrived at through the thinking process claimed to be measured not from clues or faulty assumptions. Likewise wrong answers are arrived at through faulty reasoning and not due to good reasoning based on a different philosophical orientation or experience level. This information would most likely be obtained by observing or interviewing students.

c. There is a moderate correlation with intelligence and achievement tests. Scores correlate with other validated tests claiming to measure the same thing.

d. There is a factor analysis done to show that the subscales do measure different things.

e. Groups that should be different in their scores are indeed different. This could include the ability of an instrument to differentiate between types of students.

f. The instrument measures changes or differences in HOTS after training designed to change HOTS.

g. There is a clear and frank discussion of the measurement issues involved including which aspects were investigated during the development process and which were not.

h. It is the opinion of knowledgeable judges that the instrument measures the HOTS aspects claimed.

i. For Piagetian instruments there is a high correlation between scores on the test and level of formal reasoning obtained from clinical interviews.

2. For open-ended instruments this would include:

a. The respondent understands what is being asked. Vocabulary or concepts unfamiliar to a group would make the instrument unusable for that group.

This information would most likely be obtained by observing or interviewing students.

b. There is a moderate correlation with intelligence and achievement tests. Scores from the instrument correlate with scores from other instruments claiming to measure the same thing.

c. Groups that should be different in their scores are indeed different. This could include the ability of an instrument to differentiate between types of students.

d. The instrument measures changes or differences in HOTS after training designed to change HOTS.

e. There is a clear and frank discussion of the measurement issues involved including which aspects were investigated during the development process and which were not.

f. It is the opinion of knowledgeable judges that the instrument measures the HOTS aspects claimed.

TITLE INDEX

Applications of Generalizations Test, 31
Arlin Test of Formal Reasoning, 32
Assessment of Reading Growth, 38,43
California Achievement Test, Forms E and F, 38
Classroom Test of Formal Reasoning, 35
Comprehensive Tests of Basic Skills, Forms U and V, 38
Cornell Class Reasoning Test, Form X, 15
Cornell Conditional Reasoning Test, Form X, 16
Cornell Critical Thinking Tests, Level X and, 17
Critical Thinking Test, 43
Developing Cognitive Abilities Test, 39
Educational Quality Assessment, 43
Ennis-Weir Critical Thinking Essay Test, 18
Florida Taxonomy of Cognitive Behavior, 45
Formal Operations Measure, 35
Formal Operations Test (Biology, History and Literature), 35
Group Assessment of Logical Thinking, 35
Iowa Test of Basic Skills: Early Primary and Primary Batteries, 38
Judgment: Deductive Logic and Assumption, 19
Make A Tree, 36
Means-Ends Problem Solving, 20
Metropolitan Achievement Tests, 6th edition, Form L, 38
National Tests of Basic Skills, 38
New Jersey Test of Reasoning Skills, Form B, 21
Pennsylvania Assessment of Creative Tendency, 36
Possible Jobs, 36
Primary Test of Higher Processes Thinking, 31
Purdue Elementary Problem Solving Inventory, 23
Reading Yardsticks, 38
Ross Test of Higher Cognitive Processes, 24
Scan-Tron Reading Tests, 38
Science Reasoning Level Test, 35
Seeing Problems, 36
Springs Task, 35
SRA Achievement Series, 39
Stallings-Simon Observation Instrument, 45
Stanford Achievement Test, Forms E and F, 39
Stanford Test of Academic Skills, Forms E and F, 39
Structure of Intellect Learning Abilities Test (SOI-LA), 39
Survey of Academic Skills, 45
Survey of Basic Skills, Forms P and Q, 39
TAB Science Test: An Inventory of Science Methods, 31
Test of Creative Potential, 36
Test of Divergent Feeling, 36
Test of Divergent Thinking, 36
Test of Enquiry Skills, 31
Test of Logical Thinking, 35
Test of Science Comprehension, 31
Test on Appraising Observations, 26
Think It Through, 28
Thinking Creatively With Sounds and Words, 37
Thinking Skills Teaching Inventory, 45
Three-R's Tests, Forms A and B, 39
Torrance Test of Creative Thinking, 37
Understanding in Science, 34
Valett Inventory of Critical Thinking Abilities, 35
Watson-Glaser Critical Thinking Appraisal, 29
Williams Scale, 36

INDEX OF TESTS BY GRADE LEVEL

Elementary (K-6)

Applications of Generalizations Test (4-12), 31
Arlin Test of Formal Reasoning (6-Adult), 32
Assessment of Reading Growth (3,7,11), 38,43
California Achievement Test, Forms E and F (K-12), 38
Comprehensive Tests of Basic Skills, Forms U and V (K-12), 38
Cornell Class Reasoning Test, Form X (4-12), 15
Cornell Conditional Reasoning Test, Form X (4-12), 16
Cornell Critical Thinking Tests, Level X, 17
Developing Cognitive Abilities Test (2-12), 39
Group Assessment of Logical Thinking (6-12), 35
Iowa Test of Basic Skills: Early Primary and Primary Batteries (K-12), 38
Make A Tree (PreK-1), 36
Means-Ends Problem Solving (5-7), 20
Metropolitan Achievement Tests, 6th edition, Form L (K-12), 38
National Tests of Basic Skills (PreK-College), 38
New Jersey Test of Reasoning Skills, Form B (4-College), 21
Pennsylvania Assessment of Creative Tendency (4-9), 36
Possible Jobs (6-12), 36
Primary Test of Higher Processes Thinking (2-4), 31
Purdue Elementary Problem Solving Inventory (2-6), 23
Reading Yardsticks (K-8), 38
Ross Test of Higher Cognitive Processes (4-6), 24
Scan-Tron Reading Tests (3-8), 38
Science Reasoning Level Test (3-6), 35
Springs Task (5-Adult), 35
SRA Achievement Series (K-12), 39
Stanford Achievement Test, Forms E and F (1-9), 39
Structure of Intellect Learning Abilities Test (SOI-LA) (K-Adult), 39
Survey of Basic Skills, Forms P and Q (K-12), 39
TAB Science Test: An Inventory of Science Methods (4-6), 31
Test of Creative Potential (2-12), 36
Test of Divergent Feeling (1-12), 36
Test of Divergent Thinking (1-12), 36
Test of Logical Thinking (6-College), 35
Test of Science Comprehension (4-6, upper elementary), 31
Think It Through (4-6), 28
Thinking Creatively With Sounds and Words (2-Adult), 37
Three-R's Tests, Forms A and B (K-12), 39
Torrance Test of Creative Thinking (K-Adult), 37
Valett Inventory of Critical Thinking Abilities (PreK-6), 35
Williams Scale (1-12), 36

Junior High (7-8)

Applications of Generalizations Test (4-12), 31
Arlin Test of Formal Reasoning (6-Adult), 32
Assessment of Reading Growth (3,7,11), 38,43
California Achievement Test, Forms E and F (K-12), 38
Classroom Test of Formal Reasoning (8-12), 35
Comprehensive Tests of Basic Skills, Forms U and V (K-12), 38

Cornell Class Reasoning Test, Form X (4-12), 15
Cornell Conditional Reasoning Test, Form X (4-12), 16
Cornell Critical Thinking Tests, Level X, 17
Developing Cognitive Abilities Test (2-12), 39
Formal Operations Test (Biology, History & Literature) (8-Adult), 35
Group Assessment of Logical Thinking (6-12), 35
Iowa Test of Basic Skills: Early Primary and Primary Batteries (K-12), 38
Judgment: Deductive Logic and Assumption Recognition (7-12), 19
Means-Ends Problem Solving (5-7), 20
Metropolitan Achievement Tests, 6th edition, Form L (K-12), 38
National Tests of Basic Skills (PreK-College), 38
New Jersey Test of Reasoning Skills, Form B (4-College), 21
Pennsylvania Assessment of Creative Tendency (4-9), 36
Possible Jobs (6-12), 36
Reading Yardsticks (K-8), 38
Ross Test of Higher Cognitive Processes (4-6), 24
Scan-Tron Reading Tests (3-8), 38
Seeing Problems (7-Adult), 36
Springs Task (5-Adult), 35
SRA Achievement Series (K-12), 39
Stanford Achievement Test, Forms E and F (1-9), 39
Stanford Test of Academic Skills, Forms E and F (8-13), 39
Structure of Intellect Learning Abilities Test (SOI-LA) (K-Adult), 39
Survey of Basic Skills, Forms P and Q (K-12), 39
Test of Creative Potential (2-12), 36
Test of Divergent Feeling (1-12), 36
Test of Divergent Thinking (1-12), 36
Test of Enquiring Skills (7-10), 31
Test of Logical Thinking (6-College), 35
Test on Appraising Observations (7-Adult), 26
Thinking Creatively With Sounds and Words (2-Adult), 37
Three-R's Tests, Forms A and B (K-12), 39
Torrance Test of Creative Thinking (K-Adult), 37
Understanding in Science (7-9), 34
Valett Inventory of Critical Thinking Abilities (PreK-6), 35
Williams Scale (1-12), 36

<u>High School (9-12)</u>

Applications of Generalizations Test (4-12), 31
Arlin Test of Formal Reasoning (6-Adult), 32
Assessment of Reading Growth (3,7,11), 38,43
California Achievement Test, Forms E and F (K-12), 38
Classroom Test of Formal Reasoning (8-12), 35
Comprehensive Tests of Basic Skills, Forms U and V (K-12), 38
Cornell Class Reasoning Test, Form X (4-12), 15
Cornell Conditional Reasoning Test, Form X (4-12), 16
Cornell Critical Thinking Tests, Level X, 17
Developing Cognitive Abilities Test (2-12), 39
Ennis-Weir Critical Thinking Essay Test (9-Adult), 18
Formal Operations Test (Biology, History & Literature) (8-Adult), 35
Group Assessment of Logical Thinking (6-12), 35
Iowa Test of Basic Skills: Early Primary and Primary Batteries (K-12), 38
Judgment: Deductive Logic and Assumption Recognition (7-12), 19

Means-Ends Problem Solving (5-7), 20
Metropolitan Achievement Tests, 6th edition, Form L (K-12), 38
National Tests of Basic Skills (PreK-College), 38
New Jersey Test of Reasoning Skills, Form B (4-College), 21
Pennsylvania Assessment of Creative Tendency (4-9), 36
Possible Jobs (6-12), 36
Seeing Problems (7-Adult), 36
Springs Task (5-Adult), 35
SRA Achievement Series (K-12), 39
Stanford Achievement Test, Forms E and F (1-9), 39
Stanford Test of Academic Skills, Forms E and F (8-13), 39
Structure of Intellect Learning Abilities Test (SOI-LA) (K-Adult), 39
Survey of Basic Skills, Forms P and Q (K-12), 39
Test of Creative Potential (2-12), 36
Test of Divergent Feeling (1-12), 36
Test of Divergent Thinking (1-12), 36
Test of Enquiring Skills (7-10), 31
Test of Logical Thinking (6-College), 35
Test on Appraising Observations (7-Adult), 26
Thinking Creatively With Sounds and Words (2-Adult), 37
Three-R's Tests, Forms A and B (K-12), 39
Torrance Test of Creative Thinking (K-Adult), 37
Understanding in Science (7-9), 34
Watson-Glaser Critical Thinking Appraisal (9-Adult), 29
Williams Scale (1-12), 36

College/Adult

Arlin Test of Formal Reasoning (6-Adult), 32
Cornell Critical Thinking Tests, Level Z, 17
Ennis-Weir Critical Thinking Essay Test (9-Adult), 18
Formal Operations Test (Biology, History & Literature) (8-Adult), 35
Formal Operations Measure (College), 35
Means-Ends Problem Solving (5-7), 20
National Tests of Basic Skills (PreK-College), 38
New Jersey Test of Reasoning Skills, Form B (4-College), 21
Seeing Problems (7-Adult), 36
Springs Task (5-Adult), 35
Stanford Achievement Test, Forms E and F (1-9), 39
Stanford Test of Academic Skills, Forms E and F (8-13), 39
Test of Logical Thinking (6-College), 35
Test on Appraising Observations (7-Adult), 26
Thinking Creatively With Sounds and Words (2-Adult), 37
Torrance Test of Creative Thinking (K-Adult), 37
Watson-Glaser Critical Thinking Appraisal (9-Adult), 29

THE TEST CENTER

The Test Center at the Northwest Regional Educational Laboratory is a library of tests and testing resources. Materials are loaned to educators in Alaska, Hawaii, Idaho, Montana, Oregon, Washington and the Pacific Islands; and to Chapter 1 programs in Arizona, California, Colorado, New Mexico, Nevada, Utah and Wyoming. Most of the Higher Order Thinking Skills tests in this guide are available for a three week load by contacting:

The Test Center
Northwest Regional Educational Laboratory
101 SW Main Street, Suite 500
Portland, OR 97204
503/275-9500 or 800/547-6339